AD615324

$7.40

QUARTERLY PROGRESS REPORT

No. 77

APRIL 15, 1965

DDC

MAY 26 1965

TISIA B
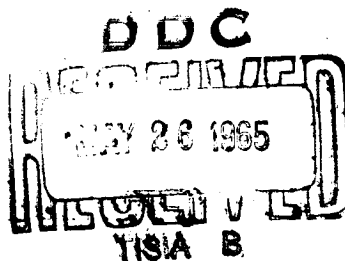
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
CAMBRIDGE, MASSACHUSETTS

GPO PRICE $ _____

CSFTI PRICE(S) $ _____

Hard copy (HC) ___7.36___

Microfiche (MF) ___2.25___

ff 653 July 65

Acquisitioned Document
SQT

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

QUARTERLY PROGRESS REPORT No. 77

April 15, 1965

Submitted by: H. J. Zimmermann
G. G. Harvey

TABLE OF CONTENTS

CONTENTS

√ PLASMA DYNAMICS

CONTENTS

COMMUNICATION SCIENCES AND ENGINEERING

CONTENTS

PERSONNEL

Administration

Professor H. J. Zimmermann, Director
Professor G. G. Harvey, Associate Director
Mr. R. A. Sayers, Assistant Director

Advisory Committee

Dean G. S. Brown
Prof. W. W. Buechner
Prof. W. B. Davenport, Jr.
Prof. P. Elias
Prof. G. G. Harvey
Prof. A. G. Hill
Prof. I. W. Sizer
Dean J. B. Wiesner
Prof. H. J. Zimmermann
    (Chairman)

Research Committee

Prof. S. C. Brown
Prof. L. J. Chu
Prof. M. Halle
Prof. G. G. Harvey
Prof. W. A. Rosenblith
Mr. R. A. Sayers
Prof. W. M. Siebert
Prof. L. D. Smullin
Prof. M. W. P. Strandberg
Prof. P. D. Wall
Prof. J. R. Zacharias
Prof. H. J. Zimmermann
    (Chairman)

Professors

Allis, W. P.	Hill, A. G.	Siebert, W. M.
Bitter, F.	Huffman, D. A.	Smullin, L. D.
Brown, S. C.	Jakobson, R.	Stevens, K. N.
Chomsky, A. N. (Absent)	Kurylowicz, J. (Visiting)	Strandberg, M. W. P.
Chu, L. J.	Lee, Y. W.	Sutherland, N. S. (Visiting)
Eden, M.	Mason, S. J.	Wall, P. D.
Edgerton, H. E.	Minsky, M. L. (Absent)	Warren, B. E.
Elias, P.	Putnam, H.	Waugh, J. S.
Halle, M.	Rose, D. J.	Wozencraft, J. M.
Harvey, G. G.	Rosenblith, W. A.	Zacharias, J. R.
Haus, H. A.	Shannon, C. E.	Zimmermann, H. J.
	Shapiro, A. H.	

Associate Professors

Barrett, A. H.	Hammes, G. G.	McCune, J. E.
Bekefi, G.	Hoffman, M. A.	Oates, G. C.
Bers, A.	Ingard, K. U.	Peake, W. T.
Bolz, G. (Visiting)	Jackson, W. D.	Penfield, P. L., Jr.
Bose, A. G.	Jacobs, I. M. (Absent)	Pomorska, Krystyna
Brown, G. A.	Kerrebrock, J. L.	Schreiber, W. F.
Gallager, R. G.	King, J. G.	Searle, C. L.
Garland, C. W.	Kyhl, R. L.	Taylor, E. F.
Graham, J. W.	Loewenthal, M.	Teager, H. M.
Gyftopoulos, E. P.	Matthews, G. H.	Whitehouse, D. R.

vii

PERSONNEL

Assistant Professors

Algazi, V. R. (1)
Billman, K. W.
Bobrow, D. G. (1)
Brown, J. E. (Visiting)
Bruce, J. D. (1)
Carabateas, E. N. (Absent)
Colaclides, P. (Visiting)
Cooper, R. S. (1)
Dean, L. W. III
Dennis, J. B.
Dupree, T. H.
Fiocco, G.
Fodor, J. A. (Absent)
Getty, W. D.

Hall, J. L. II (Absent)
Heinz, J. M.
Heiser, W. H.
Hennie, F. C. III
Huang, T. S. (1)
Katz, J. J. (Absent)
Kennedy, R. S. (1)
Klima, E. S.
Lee, H. B. (1)
Lidsky, L. M.
MacDonald, J. S. (1)
Oppenheim, A. V. (1)
Perry, C. H.

Pfeiffer, R. R. (1)
Postal, P. M.
Rafuse, R. P.
Schetzen, M.
Schwab, W. C.
Shavit, A.
Stickney, R. E.
Stiglitz, I. C. (2) (Absent)
Stockham, T. G., Jr.
Tretiak, O. J. (1)
Troxel, D. E.
Van Trees, H. L., Jr.
Weiss, R.
Weiss, T. F. (1)

Lecturer

Ferretti, E.

Instructors

Alter, R.
Andersen, J.
Bernard, G. D.
Crystal, T. H.
Gray, P. R.
Huibonhoa, R.
Ingraham, J. C.

Kahn, R. E. (1)
Katona, P. G.
Kincaid, T. G.
Kliman, G. B.
Landsman, E. E.
Lenoir, W. B.
Parente, R. B.
Parker, R. R.

Perlmutter, D. M.
Pierson, E. S.
Sachs, M. B.
Schneider, H. M.
Spann, R. N.
Staelin, D. H.
Zeiger, H. P. (1)

Research Associates

Barnett, G. O.
Blum, M.
Cerrillo, M. V.
Dupress, J. K.
Durlach, N. I.
Edmonds, P. H.
Gesteland, R. C.

Hall, R. D.
Halverson, W. D.
Heywood, J. B.
Keyser, S. J.
Kolers, P. A.
Kornacker, K.
Lettvin, J. Y.
O'Neil, W. A.

Papert, S. A.
Rugheimer, J. H.
Smith, T. G., Jr.
Stark, L. (Absent)
Thompson, E.
Yngve, V. H.
Zisk, S. H.

Guests

Fraser, J. B.
Lakoff, G. P.
Szepe, G.

Visiting Scholar

Yamaguchi, Y.

(1) Ford Foundation Fellow   (2) Lincoln Laboratory Staff Associate

PERSONNEL

Research Affiliates

Barlow, J. S.
Brodey, W. M.
Brown, R. M.
Crist, A. H.
Cunningham, J. E.

Fohl, T.
Hart, R. W.
Howland, B.
Kaye, J. D.
Langbein, D.
Littleboy, H. S.

McLardy, T.
Molnar, C. E.
Slawson, A. W.
Walker, D. E.
Warotamisikkhadit, U.

Postdoctoral Fellows

Cunningham, A. W. B. (1)
French, T. C. (1)
Frenk, S. (2)

Lampis, G. (3)
Newmark, R. A. (4)
Pickard, Barbara C. G. (1)
Pickard, W. F. (1)

Shofer, R. J. (1)
Spivey, H. O. (1)
Taub, A. (1)

R.L.E. Research Staff

Andrews, J. M.
Badessa, R. S.
Benhaim, N.
Bosche, Carol M.
Breon, R. K.
Carter, R. J.
Charney, Elinor K.
Clayton, R. J.
Crowther, Patricia P.
Darlington, J. L.
Dolan, J. M.
Edwards, D. J.
Fabry, R. S.
Fratar, Gail M.
Geick, R.

Goodall, M. C.
Ingersoll, J. G.
Ingham, K. R.
Jordan, F. N.
Kannel, Muriel S.
Kerllenevich, N.
Kiang, N. Y-s.
Kierstead, J. D.
Levy, Rachel J.
Lewis, Ellie Reed
Lontai, Laszlo N.
McCarthy, J. J.
McCulloch, W. S.
Menyuk, Paula
Mulligan, W. J.
Musha, T.

O'Brien, F. J.
O'Rourke, Ann M.
Palfy, M. A.
Pennell, Martha M.
Pitts, W. H.
Rappaport, R. L.
River, Eleanor C.
Rojas Carona, R. R.
Rosebury, F.
Ryan, L. W.
Urbanek, K.
Vidale, Eda B.
Viertel, J. J.
Willke, H. L., Jr.
Zatorski, R. J.

Research Assistants

Arnstein, D. S. (5)
Austin, M. E. (5)
Bartsch, R. R.
Bauer, R. F.
Breeding, R. J.
Bush, A. M.
Chandra, A. N.
Chapin, P. G. (6)
Chapman, J. C.
Chase, D.
Clarke, J. F.
Clemens, J. K.

Coggins, J. L.
Dethlefsen, R.
DeWolf, J. B.
Donohew, J. N., Jr.
Edwards, K. R.
Engelmaier, W.
Erman, J. E.
Falconer, D. D.
Fehrs, D. L.
Fessenden, T. J.
Froeschle, T. A.
Gabrielian, A.

Gadzuk, J. W.
Geis, M. L.
Glantz, R. S.
Goodman, L. M.
Grams, G. W.
Gustafson, K. T.
Harlem, G. S.
Harris, J. W.
Hill, R. A.
Hofmann, T. R.
Huang, T.
Jameson, P. W.

(1) National Institutes of Health Fellow
(2) Rockefeller Foundation Fellow
(3) Senior Fellow, NASA

(4) National Science Foundation Fellow
(5) Lincoln Laboratory Staff Associate
(6) Danforth Fellow

Research Assistants (continued)

Kalelkar, A. S.
Kapetanakos, C. A.
Keating, R. F.
Kilham, L. B.
King, R. W.
Kiparsky, R. P. V.
Kniazzeh, A. G. F.
Koons, H. C.
Kusse, B. R.
Kyong, S. H.
Landers, Elisabeth K.
Lightner, T. M.
Lind, M. M.
Liu, J-H.
Logan, R. M.
Lubin, B. T.
Lutz, M. A.
Mark, R. G.
Maul, M. K.
Max, J.
McNary, C. A.

Mendelsohn, R. L.
Milne, D. C.
Molden, J. C.
Moran, J. M.
Moses, J.
Nahvi, M.
Nelsen, D. E.
Offenberger, A. A.
Pauwels, H. J. E. H.
Peters, P. S., Jr.
Porter, R. P.
Portner, E. M., Jr.
Posner, R. D.
Poussart, D. J. M.
Pruslin, D. H.
Qualls, C. B.
Renard, R. H.
Ribbeck, C. S.
Rogers, A. E. E.
Ross, J. R.
Simon, E. M.
Snyder, D. L.

Solbes, A.
Spiridon, A.
Spoor, J. H.
Steinbrecher, D. H.
Sun, P. B-S.
Sutherland, W. R. (1)
Taylor, M. G.
Thome, R. J.
Tomlinson, R. S.
Wagner, C. E.
Wang, C. H.
Warshawsky, Florence
Wiederhold, M. L.
Williams, J. A.
Williams, J. T.
Winett, J. M. (1)
Wissmiller, J. C.
Woo, J. C.
Yamamoto, S.
Yarnell, C. F.
Zeiders, G. W., Jr.

Graduate Assistants

Allen, R. J.
Andrews, M. L.
Brenner, F. J.
Fukumoto, A.
Garosi, G. A.
George, E. V.
Golub, R.
Guttrich, G. L.
Hooper, E. B., Jr.

Johnston, W. D., Jr.
Katyl, R. H.
Kronquist, R. L.
Kukolich, S. G.
Lewis, T. B.
Llewellyn-Jones, D. T.
Macon, J. L.
Manheimer, W. M.
Mattison, E. M.

Pleasance, L. D.
Reifenstein, E. C. III
Reznek, S. R.
Rogoff, G. L.
Rosen, L.
Swain, D. W.
Tse, F. Y-F.
Waletzko, J. A.
Young, E. F.

Teaching Assistants

Burns, S. K.
Cornew, R. W.
Donaldson, R. W.
Graham, D. N.
Guttman, D. S.
Hamawi, J. N.
Henke, W. L.
Leonardi-Cattolica, A. M.
Levin, M. I.

Lou, D. Y-S.
Marcus, P. L.
Metz, P. J.III
Meyn, J. H.
Murray, M. J.
O'Lague, P. H.
O'Leary, G. C.
Poulo, L. R.
Prabhu, V. K.

Robertson, E. A.
Samis, M. H.
Schaefer, D. W.
Schane, S. A.
Smith, T. B.
Speck, C. E.
Stone, E. T.
Thompson, D. S.
Veneklasen, L. H.

(1) Lincoln Staff Associate

PERSONNEL

Graduate Students

Acker, D. E. (1)
Anderson, J. A. (2)
Arnold, C. E.
Baggeroer, A. B. (3)
Baker, T. H.
Bever, T. G.
Braida, L. D. (3)
Brown, T. S.
Caldwell, D. (3)
Carpenter, R. A. (3)
Cesarski, W. W. (4)
Chan, S. W-C. (5)
Cohen, A. J. (3)
Collins, L. D. (3)
Colton, J. R. (6)
Davis, E. S.
Davis, J. A. (3)
Doane, J. L. (7)
Domen, J. K.
Drechsler, R. C. (6)
Dunn, F. E. (3)
Ebert, P. M. (3)
Evans, S. A. (4)
Fetz, E. E. (3)
Fidelholtz, J. L. (8)
Flannery, D. L. (3)
Fleury, P. A. (3)
Flynn, R. W. (4)
Foley, J. A., Jr. (8)
Freeman, J. A.
Gaut, N. E. (9)
Gentle, K. W. (3)
Gerry, E. T.
Glaser, J. (3)
Glenn, W. H., Jr.

Gruber, J. S. (8)
Guinan, J. J., Jr. (3)
Hall, Barbara C. (3)
Hartline, D. K. (10)
Hebel, W. T., Jr.
Hoffman, M. A. (6)
Hsi, C-F. G. (4)
Ishii, H. (11)
Jenkins, L. (3)
Johnson, C. J. (6)
Kinzer, T. J. III (8)
Krakauer, L. J. (3)
Kuroda, S-Y. (12)
Lea, W. A. (3)
Lee, F. F. (7)
Leech, G. N. (13)
Lercari, R. F. (14)
Levin, B. J. (6)
Lieberman, M. A. (3)
Lowell, F. C., Jr.
Lubin, M. D.
Lynch, J. T. (14)
Maduemezia, A. A. (15)
Maimon, S. R. (14)
Martinelli, M. A. (3)
McCann, R. J. (6)
McDowell, G. Q. (3)
McEnally, T. E., Jr. (16)
McMorris, J. A. II
Mendell, L. M. (3)
Merrill, E. G. (2)
Moir, R. W. (4)
Mozzi, R. L.
Niessen, C. W. (3)
Nolan, J. J., Jr.
Nyman, T. H. (6)

Pilc, R. (6)
Pinkston, J. T. III (3)
Portinari, J. C. (17)
Rabiner, L. R. (3)
Rezende, S. M. (18)
Richters, J. S. (3)
Riehl, J. W. (19)
Rosenbaum, P. S. (8)
Ross, J. A. (3)
Scholl, M. M. (20)
Schulz, H. M. III (3)
Scoville, F. W. (6)
Selleslaugh, E. (21)
Singer, J. J. (3)
Smith, C. V., Jr. (22)
Smith, R. S. (3)
Stanley, R. J. (3)
Strahm, N. D. (3)
Thomae, I. H. (3)
Thornburg, C. O., Jr.
Tomlinson, Karen E. (3)
Tse, A. N.
Turner, L. D. (14)
Vander Vorst, A. (23)
von Bismark, G. (11)
Wallace, R. N. (3)
Ward, E. D. (3)
Wawzonek, J. J.
Weiner, S. D. (3)
Wickelgren, Barbara G. (3)
Wilkins, D. R. (24)
Wolaver, D. H. (3)
Wright, B. L. (3)
Zeidenbergs, J. (25)
Zwicky, A. M., Jr. (8)

(1) Raytheon Fellow
(2) Public Health Service Trainee
(3) National Science Foundation Fellow
(4) Atomic Energy Commission Fellow
(5) Edward Austin Fellow
(6) Bell Telephone Laboratories Fellow
(7) Hertz Foundation Fellow
(8) National Defense Education Act Fellow
(9) NASA Fellow
(10) National Institutes of Health Fellow
(11) Fulbright Fellow
(12) IBM Fellow
(13) Harkness Fellow

(14) National Institutes of Health Trainee
(15) University of Nigeria Fellow
(16) Ford Foundation Fellow
(17) National Research Council Fellow
(18) Institute of International Education Fellow
(19) Proctor and Gamble Fellow
(20) Sloan Teaching Fellow
(21) Belgian-American Education Foundation Fellow
(22) Gerard Swope Fellow
(23) NATO Fellow
(24) Space Technology Laboratories Fellow
(25) Sperry Industrial Fellow

PERSONNEL

Undergraduates (Thesis or Special Problems)

Anderson, C. R.
Athans, D. P.
Bass, W. L.
Bird, P. A.
Breedlove, J. R., Jr.
Broadley, W. H.
Bromberg, S. N.
Burton, R. K.
Carr, Patricia D.
Chang, W. Y.
Cheng, J.
Cicerone, R. J.
DiGrazia, R. J.
Downward, J. G., IV
Dyro, J. F.
Eberbach, S. J.
Emo Capodilista, G.
Gill, G. Y.
Giovachino, D. L.
Graham, M. C.
Graham, M. M.
Graves, R. E., Jr.
Guldi, R. L.
Hakel, W. J.
Hall, D. B.

Harris, R. V. III
Harrison, D. C.
Huntington, W. R.
Hurd, C. F.
Jacobs, E. P.
Kampmann, E. C.
King, T. R.
Koolish, R. M.
Kurton, S.
Lake, L. R.
Litke, J. D.
Lowry, B. L.
Lubitz, P.
MacLeod, D. J.
Marino, M. J.
Maskasky, J. L.
McConchie, A. L., Jr.
McMillin, R. S.
Mechler, D. H.
Memishian, J., Jr.
Miller, D. L.
Mosier, W. E., Jr.
Mower, H. W.
Moxon, E. C.
Newton, P. E.
Nordman, B. J., Jr.

Onorato, J. M.
Pearson, D. J.
Peusner, L. J.
Phelps, J. O. III
Preer, R. M., Jr.
Prout, F. C.
Riezenman, M. J.
Sacks, S.
Seelig, W. D.
Sherman, R. D.
Shook, G. K. II
Shulman, D. L.
Siemens, P. J.
Smith, K. A.
Sramek, R. A.
Stark, L. A.
Stebbins, D. H.
Stein, M. A.
Stowell, W. R.
Weisgerber, W. R.
Wessler, B. D.
Wilensky, R.
Williams, F. K.
Wolf, J. E.
Zelazo, R. E.

Student Employees

Ackerman, W. B.
Barnwell, T. P. III
Boxman, R. L.
Brody, W. R.
Chang, K.
Crate, S. M.
Currano, J. J.
Engle, R. H.
Granek, H.

Haney, D. L.
Himka, R. L.
Howatt, J. R.
Jensen, E. R.
Koralek, R. W.
Masson, R. K.
Ogan, K. L.
Perrolle, P. M.
Rose, S. M.
Ryder, J. K.

Schwartskopf, L. A.
Seitz, C. L.
Smith, D. P.
Spitzer, S. M.
Stampfel, J. P., Jr.
Strause, P. E.
Tweed, D. G.
Ward, S. A.
Westerfeld, E. C.

R.L.E. Administration Staff

Duffy, D. R.
Hewitt, J. H.
Keyes, R. V., Jr.

Sayers, R. A.
Smith, P. L.
Thomas, Helen L.

Administrative Assistant

Bella, C. J.

Office Clerks

Barron, Gladys G.
Billings, W. J.
Chaudari, Karin
Engler, R. R.

Goody, Elfriede
Gregor, C. A.
Peck, J. S.

Ruggere, P. A.
Scalleri, Mary B.
Stagliori, Eleanor E.
Toebes, Rita K.

PERSONNEL

Typists

Beane, Mildred E.
Findley, Janet P.
Foley, Ruth E.

Mackenzie, Jean C.
Murphy, Mary R.
Myers, Alberta L.

Technical Typists

Barnes, R. A.
Capron, Evelyn L.

Fleming, Patricia L.
Smee, P. E.

Secretaries

Aberle, Carol A.
Bella, Rose Carol
Boucher, Madeleine I.
Carbone, Angelina
Cohen, Phyllis J.
Conwicke, Vera
Cummings, Jane F.
Finer, Faith L.
Geller, Elaine J.
Gordon, Linda S.
Hamilton, Martha C.
Iverson, Alice I.

Johnson, Barbara A.
Kaloyanides, Venetia
Koontz, Linda E.
Lannoy, Doris E.
Laurendeau, Carole A.
Loeb, Charlotte G.
McCarthy, Barbara L.
McEntee, Doris C.
Morneault, Beverly A.
Murray, Maureen P.
Notman, Mary S.
Owens, Mary E.
Parenteau, Joan C.

Parrella, Cynthia A.
Petone, Rosina C.
Pierce, Marilyn A.
Ponte, Mary J.
Ricker, Barbara J.
Romanos, Eugenia A.
Russell, Susan P.
Smith, Clare F.
Smythe, Carol A.
Taylor, Anne L.
Thorne, Jutta
Wanner, Patricia A.

Engineering Assistants

Barrett, J. W.
Berg, A. E.

Crist, F. X.
Fontaine, C. L.
McKenzie, J. A.

Papa, D. C.
Thompson, J. B.

Technical Assistants

Arnn, JoAnn
Blum, G.
Byers, F. H.
Chase, Arbella P.
Grande, Esther D.

Heinz, Jane A.
Hull, A. B.
Major, Diane
Rabin, Sylvia G.
Rosenthal, Kathryn F.

Shipley, Jenot W.
Swenson, Judith E.
T,sou, B. K-Y.
Westez, C. A. H.
Yaffee, M. A.

Technicians (Assigned to Research Groups)

Babcock, E. B.
Barrows, F. W.
Bauer, T. K.
Butler, R. E., Jr.
Connolly, J. T.
Cranmer, R. E.
DiPietro, P. J.
Fitzgerald, E. W., Jr.
Gay, H. D.

Kaufman, D. E.
Kelly, M. A.
Lewis, R. R.
MacDonald, J. R.
Massey, L. N.
McCarthy, P. F.
McLean, J. J.
Misail, G. S.
Neal, R. W.

North, D. K.
Owens, R. C., Jr.
Robey, H. C.
Schwabe, W. J.
Sears, A. R.
Sprague, L. E.
Stevens, J. A.
Tortolano, A. J.
Yee, F. Q.

PERSONNEL

Technicians' Shop

Lorden, G. J., Foreman
Banks, E. C.

Fownes, Marilyn R.

Hill, R. F.
MacDonald, K. R.

Laboratory Assistants

Beaton, Catherine M.
Hurley, Susan

Levine, P. H.
Miller, S. A.

Drafting Room

Navedonsky, C. P., Foreman
Donahue, J. B.

Hillier, Anna M.

Porter, Jean M.
Rollins, I. E.

Photographic Shop

Aquinde, P.
Cook, J. F.
Karas, P.

Machine Shop

Keefe, J. B., Foreman
Aalerud, R. W.
Barnet, F. J.
Bletzer, P. W.
Brennan, J.
Bunick, F. J.

Cabral, M., Jr.
Carter, C. E.
Harvey, A. O.
Liljeholm, F. H.
Muse, W. J.

Reimann, W.
Ridge, P. A.
Ryan, J. F.
Sanromá, J. B.
Shmid, E.
Wentworth, W. G., Jr.

Tube Laboratory

Rosebury, F.

Aucella, Alice G.
Griffin, J. L.
Leach, G. H., Jr.
MacDonald, A. A.

Ryan, L. W.

Glass Shop

DiGiacomo, R. M.
Doucette, W. F.

Stock Clerks

Doherty, R. H.
Haggerty, R. H.

Legier, D. O.

Pines, J. O.
Sharib, G.

Utility and Maintenance

Doiron, E. J., Foreman
Audette, A. G.

Lucas, W. G.
McDermott, J. F.
Riley, J. F.

Sincuk, J., Jr.
Thibodeau, D. S.

PUBLICATIONS AND REPORTS

MEETING PAPERS PRESENTED

National Academy of Sciences, Madison, Wisconsin
October 11-13, 1964

> D. C. Teas and N. Y-s. Kiang, Components of the Complex Response Evoked from the Auditory Cortex of Unanesthetized Cats (invited)

Lecture, Center for Dynamic Systems, Brown University, Providence, Rhode Island
December 9, 1964

> M. C. Goodall, Rapprochement between Discrete and Continuous Systems (invited)

Lecture, IEEE Combined Groups of Circuit Theory, Information Theory, and Electron Devices, L. C. Smith College of Engineering, Syracuse University, Syracuse, New York
December 10, 1964

> S. J. Mason, Sensory Aids Research (invited)

IX Interamerican Congress of Psychology, Miami, Florida
December 16-21, 1964

> P. A. Kolers, Recall of Bilingually Presented Words

Linguistic Society of America Meeting, New York
December 28, 1964

> R. Kiparsky, Analogy and Sound Change

> S. A. Schane, Cyclic Rules in the Phonological Component of a Transformational Grammar of French

Psychology Department Colloquium, Brown University, Providence, Rhode Island
January 13, 1965

> N. Y-s. Kiang, Stimulus Coding in the Auditory Nervous System: An Electro-physiological Analysis (invited)

Annual Meeting of the American Physical Society, New York
January 27-30, 1965

> A. N. Chandra and W. D. Jackson, Superconducting Solenoids under AC Excitation

Symposium on Acoustical Physics, American Physical Society Meeting, New York
January 27-30, 1965

> U. Ingard, Nonlinear Effects and Other Phenomena in Sound Fields (invited review paper)

Seminar of Neurophysiology, Faculté des Sciences, Paris, France
February 2, 1965

> W. A. Rosenblith, Nouvelles Données sur la Codage de la Stimulation dans le Système Auditif (invited)

International Solid State Circuits Conference, University of Pennsylvania, Philadelphia, Pennsylvania

February 17-19, 1965

> D. H. Steinbrecher, Analog Circuit for Determining the Ratio and Product of Two Time Functions

Seminar, Institute of Aerospace Studies, University of Toronto, Ontario, Canada

February 18, 1965

> U. Ingard, Nonlinear Effects and Other Phenomena in Sound Fields (invited)

JOURNAL ARTICLES ACCEPTED FOR PUBLICATION

> (Reprints, if available, may be obtained from the Document Room, 26-327, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139.)

V. R. Algazi, Generalized Signal-to-Noise Ratio and Performance Index of Filters for Waveform Estimation (IEEE Trans. (SET))

M. Arbib and M. Blum, Machine Dependence of Degrees of Difficulty (Proc. Am. Math. Soc.)

F. Bitter, Ultrastrong Magnetic Fields (Sci. American)

A. Cavaggioni and M. H. Goldstein, Jr., Facilitation and Inhibition in the Visual System after Photic Stimulation (Arch. ital. Biol. (Pisa))

G. Fiocco, Optical Radar Results and Ionospheric Sporadic E (J. Geophys. Res.)

R. Geick, Evidence for a Considerable Nonlinear Dipole Absorption in Gallium Arsenide (Phys. Rev.)

J. L. Hall II, A Neural Model for Binaural Localization (IEEE Student J.)

F. C. Hennie, One-tape, Off-line, Turing Machine Computations (Inform. Contr.)

T. S. Huang, The Subjective Effect of Two-dimensional Pictorial Noise (IEEE Trans. (PTGIT))

T. S. Huang and H. B. Lee, Bounds on Impedance Functions of R, $\pm$L, $\pm$C, T Networks (J. Franklin Inst.)

J. C. Ingraham and S. C. Brown, The Helium Afterglow and the Decay of the Electron Energy (Phys. Rev.)

J. J. Katz, Semantic Theory and the Meaning of 'Good' (Philos. Rev.)

Y. W. Lee and M. Schetzen, Measurement of the Wiener Kernels of a Nonlinear System by Crosscorrelation (Internatl. J. Contr.)

D. L. Morse, Plasma Rotation in a Hollow-Cathode Discharge (Phys. Fluids)

Paul Penfield, Jr., Passivity Conditions (IEEE Trans. (CT))

R. R. Pfeiffer and N. Y-s. Kiang, Patterns of Spontaneous and Continuously Stimulated Spike Discharges in the Cochlear Nucleus of Anesthetized Cats (Biophys. J.)

JOURNAL PAPERS ACCEPTED FOR PUBLICATION (continued)

D. J. Rose, General Criterion for Mirror Instability of a Plasma (Phys. Fluids)

D. J. Sakrison, A Continuous Kiefer-Wolfowitz Process (Ann. Math. Statist.)

H. L. Witting and E. P. Gyftopoulos, Ionization Processes in a Low-Energy Cesium Plasma (J. Appl. Phys.)

LETTERS TO THE EDITOR ACCEPTED FOR PUBLICATION

A. Bers and S. Gruber, Negative-Energy Plasma Waves and Instabilities at Cyclotron Harmonics (Phys. Rev.)

H. B. Lee, Bounds on the Natural Frequencies of LC Structures (IEEE Trans. (CT))

L. M. Mendell and P. D. Wall, Response of Single Dorsal Cord Cells to Peripheral Cutaneous Unmyelinated Fibres (Nature)

P. Penfield, Jr., Maximum Cutoff Frequency of Varactor Diodes (Proc. IEEE)

P. Penfield, Jr., Note on Fourier Coefficients of Power-Law Devices (Proc. IEEE)

P. Penfield, Jr. and H. A. Haus, Electrodynamics of Moving Media (Proc. IEEE)

A. W. Rowe and J. L. Kerrebrock, Nonequilibrium Electric Conductivity of Two-Phase Metal Vapors (AIAA J.)

L. D. Smullin, A Criticism of "Energetic Electrons from a Beam Plasma Overstability" (Phys. Fluids)

D. S. Thompson and J. S. Waugh, Adjustable Ruby Intensity Standard for ESR Spectra (Rev. Sci. Instr.)

TECHNICAL REPORTS PUBLISHED

(These and previously published technical reports, if available, may be obtained from the Document Room, 26-327, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139.)

432 Alan V. Oppenheim, Superposition in a Class of Nonlinear Systems

433 L. J. Chu, H. A. Haus, and P. Penfield, Jr., The Force Density in Polarizable and Magnetizable Fluids

SPECIAL PUBLICATIONS

W. D. Jackson, J. R. Ellis, Jr., and G. O. Barnett, Calibration of Electromagnetic Flowmeters for Blood-Flow Measurement (Proc. 5th International Conference on Medical Electronics, Liége, Belgium, July 22-26, 1964)

J. J. Katz, The Philosophy of Language (a chapter in Harper Guide to Philosophy, edited by A. Danto. Harper and Rowe, New York, 1965)

W. A. Rosenblith, Engineering in the Sciences of Life and Man (Listen to Leaders in Engineering, edited by A. Love and J. S. Childers. Tupper and Love, Atlanta, 1965, pp. 275-288)

INTRODUCTION

This report, the seventy-seventh in a series of quarterly progress reports issued by the Research Laboratory of Electronics, contains a review of the research activities of the Laboratory for the three-month period ending February 28, 1965. Since this is a report on work in progress, some of the results may not be final.

GENERAL PHYSICS

I. MOLECULAR BEAMS

Prof. J. R. Zacharias   R. S. Badessa   S. G. Kukolich
Prof. J. G. King   J. F. Brenner   F. J. O'Brien
Prof. C. L. Searle   R. Golub   R. D. Posner
Prof. K. W. Billman   G. L. Guttrich   C. O. Thornburg, Jr.
Prof. E. F. Taylor   W. D. Johnston, Jr. L. H. Veneklasen

A. NEUTRALITY OF ATOMS (ATOMIC BEAM METHOD)

To provide an experimental determination of an upper limit of possible charge residing on an un-ionized atom, a cesium beam experiment is in progress. In it, a beam of atoms is passed through a uniform electric field directed perpendicular to the beam's path. Hence a Cs atom would be deflected if it bore a net charge, q. The present experimental arrangement is such that a deflection of 2 Å would be expected if $q = 10^{-18}$ e = 1.6×10^{-37} coulomb.

Orginally the experiment was performed with a DC electric field in the deflecting region. This resulted in a value $q = (8.4 \pm 8.0) \times 10^{-16}$ e. Further improvements, including the rejection, to a large extent, of beam deflections that result from field gradients, vibration of the apparatus, and similar effects, have been made by modulating the deflecting field and looking for the signal appropriate to a true Cs charge with a lock-in detector. A method for in situ calibration of the apparatus sensitivity, so essential in any null experiment, by electrically deflecting the beam a known number of angstroms at frequent intervals during data accumulation, has also been incorporated. The most recent determinations have resulted in a value $q = (-2.0 \pm 1.0) \times 10^{-18}$ e. This result should be compared with the value $q = (1.3 \pm 5.6) \times 10^{-17}$ e obtained by Zorn, and his co-workers.[1]

The present result can be interpreted in terms of possible neutron, q_n, and electron-proton pair, q_{ep}, charges that give for the cesium atom $q = 55\, q_{ep} + 78\, q_n$. Assuming $q_n = q_p$ yields $q_{ep} = q_n = (-1.5 \pm 0.75) \times 10^{-20}$ e. The precision of this result approaches that obtained by gas efflux methods.[2]

K. W. Billman, J. G. King

References

1. J. C. Zorn, G. E. Chamberlain, and V. W. Hughes, Phys. Rev. 129, 6 (1963).

2. J. G. King, Quarterly Progress Report No. 76, Research Laboratory of Electronics, M.I.T., January 15, 1965, pp. 9-13.

B. LOW TEMPERATURE HELIUM BEAM EXPERIMENT

1. Detector Operation

The following changes have been made in the detector assembly (see Quarterly Progress Report No. 76 (pages 13-15).

1. The P-16 phosphor screen has been replaced with Willemite. This is slower (decay time, microseconds) but has several hundred times higher luminous efficiency for 12-kv ions.

2. The needle holder has been adapted to point the needle directly at the phosphor screen. A study of the field electron emission pattern from the needle tip has shown that almost all field lines from the sharp tip region emerge within a 15° half-angle cone of divergence from the needle shaft axis.

3. The method of mounting the Pyrex disc on which the phosphor screen is deposited has been modified. In the old design an aluminum capture ring and a 0.020-inch gold wire gasket were used under the 1/2 inch thick glass disc. This resulted in fracture of the glass disc under temperature cycling when cooling the photomultiplier tube with dry ice. In the new configuration a stainless-steel capture ring and a 0.028-inch neoprene flat gasket which are used under the glass have thus far proved completely reliable. Also, 3/4 inch thick glass discs are now being used. The use of the neoprene seal has not resulted in poorer vacuum conditions, as a detector chamber pressure of 2×10^{-10} torr is still obtained, with or without dry-ice cooling of the photomultiplier-glass disc assembly.

Initial performance data of the detector were taken by reading the photomultiplier output current with a Keithley Model 600 electrometer, and by varying the partial pressures of N_2, A, and He in the detector chamber through the range 1×10^{-7} -2×10^{-10} torr. Pressure was monitored by a Bayard-Alpert gauge, with the correction factor appropriate for the various gases applied.[1] An over-all detection conversion efficiency (number of electrons from photomultiplier photo-cathode divided by number of atoms incident on needle tip, per unit time) of from 5% to 30% was calculated.

The sensitivity for He is approximately 4 times that for N_2, which is approximately 3/2 that for A. This is to be expected, as the phosphor yield should go as the momentum of the exciting ions, hence as the square root of the mass ratios. There is a further slight enhancement, since the helium, with a higher ionization potential, must approach the tip more closely to be ionized, the He^+ ions are formed in a higher potential region than, say, N_2^+, and hence have more energy when striking the phosphor.

The signal intensity now goes as the square of the needle tip voltage. The elimination of the electroluminescent effect (intensity exponential with tip voltage) noted previously is attributed to using Willemite rather than P-16, and is consistent with the observation that zinc-silicate base phosphors are, as a class, much less electroluminescent than zinc-sulfide base phosphors, or other common phosphor types.[2]

2. Beam Investigations

The room-temperature beam source must be operated at an excessive pressure to produce a signal visible on the Keithley electrometer, and a well-defined beam could not be obtained in this way. At a proper source pressure of the order of 10^{-4} torr one would expect approximately 500 atoms per second detected beam. Accordingly, pulse-counting electronic equipment has been incorporated. The arrangement, at present, consists of an emitter follower preamplifier, a Franklin Model 348 linear amplifier with internal pulse-height selector, and an R. L. E. Model 210 pulse counter (scale of 2^{15}). The overall resolution time is 5 μsec, and is judged appropriate for the Willemite phosphor.

This equipment has just been adjusted for optimum operation, and only preliminary results are now available. The counting statistics for the background gas fit the assumption of a random distribution, that is, the standard deviation of the means agrees within 5% of the $\sqrt{N}$ expected, when allowance is made for the counter inefficiency resulting from the 5-μsec resolution time and the 0.3% random timing error of the electro-mechanical counter timing gate.

There are preliminary indications consistent with a beam of 50 counts per second, with a source pressure of 4×10^{-2} torr; however, this has not been confirmed by scanning the needle across the beam.

An electronic gate is to be added between the Franklin and the counter for use with the beam chopper in taking velocity distributions. It is to be noted that the detected beam intensity goes _inversely_ as the 3/2 power of the source temperature, so that at 0.9° K, with the vapor pressure of He^4 is 4×10^{-2} torr, the detected beam would be 2500 counts per second, even after a 0.1% duty cycle in the chopper. Thus, from a signal-to-noise point of view, this experiment will be substantially easier with low-temperature sources.

W. D. Johnston, Jr.

References

1. S. Dushman, Scientific Foundation of Vacuum Technique (John Wiley and Sons, Inc., New York, 2d edition, 1962), p. 324.

2. H. F. Ivey, Electroluminescence and Related Effects, Advances in Electronics and Electron Physics, Suppl. I (Academic Press, New York, 1963).

C. HIGH-RESOLUTION MEASUREMENTS OF THE 3-3 INVERSION TRANSITION IN AMMONIA

The two-cavity maser spectrometer described in previous reports[1] is now being used to study the hyperfine structure of the $J = 3$, $K = 3$, inversion line of NH_3. A resonance linewidth of 350 cps is obtained with a cavity separation of 115 cm. This constitutes

improvement in resolution by a factor of 20 over previous measurements,[2] and many more lines are now resolvable.

The main line ($\Delta F = \Delta F_1 = 0$) at zero magnetic field is split as shown in Fig. I-1. The calculations of Gordon[2] indicate that the main line will be split into three
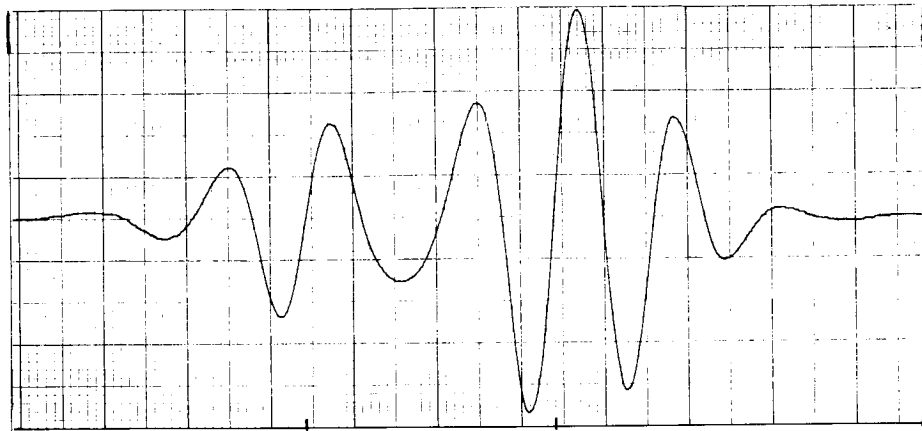
Fig. I-1.   J = 3, K = 3 main-line ($\Delta F = \Delta F_1 = 0$) components at 23,870,127.95 kc and 23,870,129.62 kc (first derivative of resonance).

components if the quadrupole coupling constant eqQ is different in the two inversion states. The separation between the $F_1 = 4$ and $F_1 = 2$ components ($\vec{F}_1 = \vec{J} + \vec{I}_N$, $I_N =$ nitrogen spin) is small (<500 cps), and the $F_1 = 2$ component should be much weaker because the degeneracy is smaller and the focussing is weaker for this state. Therefore we see primarily the $F_1 = 3$ and $F_1 = 4$ $\Delta F_1 = 0$ transitions, and the contribution from $F_1 = 2$ is much smaller. Seven of the nine components of the magnetic satellites have been observed. These magnetic satellites result when one hydrogen spin is "flipped" simultaneously with the inversion.

The main-line measurements indicate that the quadrupole coupling constant eqQ is larger in the lower inversion state by 3.6 ± 0.1 kc.

We have also observed three of the four components of one of the upper quadrupole satellites. More data will be necessary to identify each of the satellite components, but if we average over magnetic interactions we obtain agreement with Gordon's value of eqQ, 4094.8 ± 1.5 kc for the 3-3 state. When these magnetic components are identified we should obtain at least an order of magnitude improvement in accuracy for the magnetic and quadrupole coupling constants. These results provide better data on electronic wave functions and geometrical parameters for the NH_3 molecule.

S. G. Kukolich

References

1. S. G. Kukolich, Quarterly Progress Report No. 72, Research Laboratory of Electronics, M.I.T., January 15, 1964, pp. 1-11; No. 73, April 15, 1964, pp. 1-2; No. 74, July 15, 1964, pp. 4-7.

2. J. P. Gordon, Phys. Rev. 99, 1253 (1955).

II. MICROWAVE SPECTROSCOPY

Prof. M. W. Strandberg
Prof. R. L. Kyhl
Dr. J. M. Andrews, Jr.
S. N. Bromberg
G. Emo Capodilista

A. Fukumoto
M. C. Graham
R. Huibonhoa
J. G. Ingersoll
J. D. Kierstead
M. K. Maul

T. E. McEnally
R. M. Preer
S. Reznek
L. Rosen
W. J. Schwabe

A. OBSERVATIONS OF INCOHERENT PHONON PROPAGATION IN X-CUT QUARTZ

Using techniques similar to those of Gutfeld and Nethercot,[1] we have observed pulses of incoherent phonons propagating along the x axis of quartz. Figure II-1 is an oscilloscope picture of these pulses that have been generated in an aluminum film, approximately 100 Å thick, propagated along a rod of x-cut quartz, 3 mm in diameter and 19.3 mm long, and detected by means of a superconducting tin-alloy bolometer. The experiment was carried out at 3.6°K, where the maximum of the slope of the $\rho(T)$ function for the tin alloy occurred. The peak of the received signal is approximately 90 μv, corresponding to a bolometer current of 38 ma. The droop in the trailing edge of the signal at the right-hand side of the photograph is caused by the greatly decreased frequency response at 1 Mc and ringing in the video amplifier circuit. The lower trace in Fig. II-1 shows the generating pulse of ~40 watts peak power and 0.12 μsec in duration.
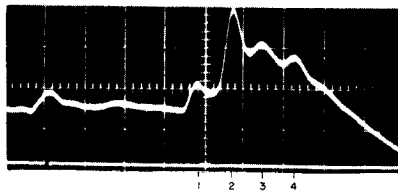
Fig. II-1.   Pulses of incoherent phonons in x-cut quartz. Lower trace shows the generating pulse which marks zero on the time axis. The oscilloscope sweep rate is 1 $\mu sec/cm$.

The symmetry of the signal pulses indicates that the mechanism responsible for the broadening is not related to thermal time constants in either the generating or the receiving metallic films.

In a previous report,[2] we pointed out that the x axis of quartz is a pure-mode axis for the longitudinal and both of the transverse elastic modes and, furthermore, that the ultrasonic Poynting vectors should be collinear with the wave vectors along this axis for each of these three modes. Thus, each of these modes should be evident in a heat pulse directed along the x axis. In Fig. II-2, we show the three modes of coherent

phonons at 0.9 Gc, generated in the usual manner,[3] by means of a re-entrant microwave cavity.
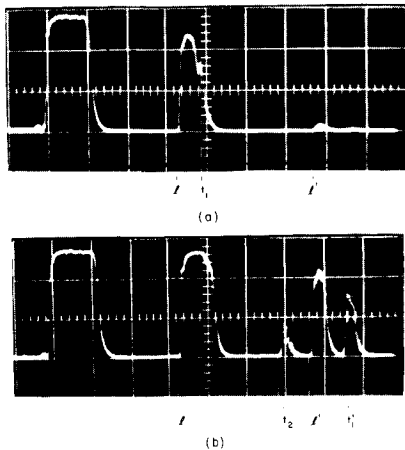
Fig. II-2.

(a) Coherent 0.9-Gc phonons in x-cut quartz. First pulse on the left is leakage from the transmitter and marks zero on the time axis.

(b) Same as (a) but with increased gain.

The oscilloscope sweep rate is 2 μsec/cm, but the pulses have made a round trip within the quartz rod and are received by the transmission cavity. Thus the scales of Fig. II-1 and II-2 can be considered the same for the purpose of total time-delay comparison. It was necessary to increase the path length in the 0.9-Gc coherent phonon experiment to resolve the longitudinal and fast-transverse modes.

In Fig. II-2a, ℓ denotes the leading edge of the first reflected longitudinal mode and t_1, the peak of the first reflected fast-transverse mode. The first reflected slow-transverse mode cannot be seen unless we increase the gain of the receiver. This is shown in Fig. II-2b, in which the peak of the slow-transverse mode is denoted t_2. Second reflections of these modes are denoted by primes in Fig. II-2. Note that, although the time scales and triggering are identical in Figs. II-2a and 2b, the leading edge of the longitudinal mode appears somewhat earlier in Fig. II-2b. This effect is caused by the finite rise time of the pulse. Thus comparisons between the leading edges of the longitudinal modes should be made between Fig. II-1 and Fig. II-2a, only.

Comparative measurements between the individual modes of the incoherent and the coherent phonon experiments are listed in Table II-1. We conclude that pulse no. 1 in the incoherent phonon experiment is actually a superposition of the longitudinal and fast-transverse pure modes. Pulse no. 4 corresponds most closely to the slow-transverse pure mode. Pulses X and Y are "extra" pulses, whose wave vectors are unknown, which happen to have Poynting vectors lying within the solid angle subtended by the receiving bolometer. In general, analytic determination of ultrasonic wave vectors as a function of the Poynting vector does not appear to be possible for an anisotropic propagation medium. Graphical methods are available,[4] and an approach by machine computation is under investigation.

At the present time, we can only speculate on the mechanism that is causing the heat pulse to broaden from 0.12 μsec to over 0.50 μsec for each of the pulses reaching the receiver. As we have pointed out in a previous report,[2] a resistive metallic film, bonded to a semi-infinite elastic medium and heated by means of a short pulse of electric energy,

Table II-1. Summary of pulse-delay data.

Experiment	Pulse	Reference	Delay (μsec)	Velocity (km/sec)	Mode
Incoherent Phonons (Fig. II-1)	1	leading edge	3.4	5.7	longitudinal
	1	peak	3.8	5.1	fast-transverse
	2	leading edge	4.3	4.5 ⎫	X
	2	peak	4.7	4.1 ⎭	
	3	leading edge	5.2	3.7 ⎫	Y
	3	peak	5.45	3.5 ⎭	
	4	leading edge	6.0	3.2 ⎫	slow-transverse
	4	peak	6.25	3.1 ⎭	
Coherent Phonons at 0.9 Gc (Fig. II-2)	ℓ	leading edge	3.35*	5.8	longitudinal
	t_1	peak	3.8*	5.1	fast-transverse
	t_2	leading edge	5.9*	3.3	slow-transverse

*Since the coherent phonon experiment involved observations of reflected pulses, actual delay times have been divided by 2 in order to correspond to the single-pass measurements made on the incoherent phonon pulses.

is capable of generating simultaneously all of the vibrational modes of an elastic propagation medium that have their Poynting vectors lying within a solid angle of 2π about any point on the surface of the film. The degree of excitation of any given mode will depend upon the characteristic of the emission of the metallic generating film and on the phonon excitation spectrum of both the generating film and the elastic propagation medium. It seems likely that the characteristic of the emission of the generating film is essentially Lambertian, modified by the effects of the elastic mismatch between the metal and the propagation medium and by the detailed geometrical nature of the interface.

Thus each of the pulses in the incoherent phonon experiment (Fig. II-1) actually contains a mixture of all of the modes whose Poynting vectors lie within the solid angle subtended by the receiving bolometer. The arrival time of each mode depends not only upon the phase velocity of the mode but also upon the angle ψ between the wave vector $\vec{k}$ and the Poynting vector $\vec{S}$ and the angle σ between the Poynting vector $\vec{S}$ and the x axis of the quartz rod. These angles are illustrated in Fig. II-3 in which we have indicated a ray of elastic energy emanating from an arbitrary point O in the metallic generating film and propagating in the direction of the Poynting vector $\vec{S}$. The cross-hatching in the ray represents surfaces of constant phase, which are perpendicular to the wave vector $\vec{k}$. It must be emphasized that Fig. II-3 is a diagram presented for the sole purpose of defining angles. Physically, a single ray of incoherent elastic energy
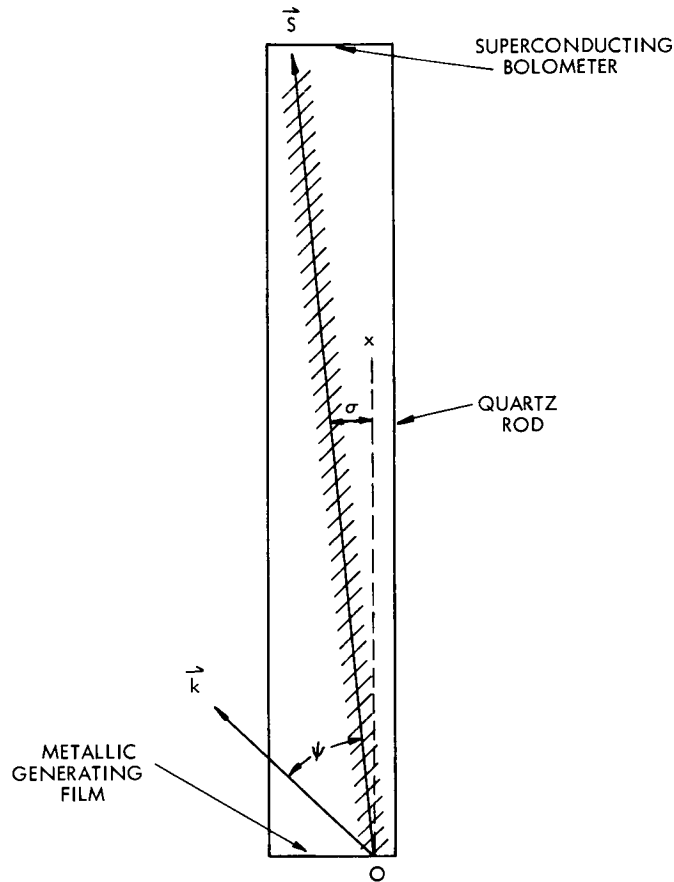
Fig. II-3. Phonon propagation in a rod of an anisotropic solid
for arbitrary wave vector $\vec{k}$. The Poynting vector
$\vec{S}$ denotes the direction of energy propagation.

cannot be characterized by a single wave vector, as there are no surfaces of constant

phase. Rather, each ray contains an infinite distribution of $\vec{k}$ vectors, varying both in

length and direction. It is just this sort of distribution in k-space that could give rise

to the pulse broadening that is evident in Fig. II-1. In order to see this more clearly,

we write an expression for the arrival time of elastic energy as a function of the wave

vector $\vec{k}$ and the angles ψ and σ, which are also functions of the wave vector $\vec{k}$.

$$T_k = \frac{\ell \cos \psi}{v_k \cos \sigma} , \tag{1}$$

where T_k is the delay time of the kth mode, v_k is the phase velocity of the kth mode,
$\psi = \psi(\vec{k})$ is the angle between wave vector $\vec{k}$ and Poynting vector $\vec{S}$, $\sigma = \sigma(\vec{k})$ is the angle
between Poynting vector $\vec{S}$ and quartz rod axis, and ℓ is the length of the quartz rod.
Thus the broadening of the incoherent pulses could arise from the distribution in the

arrival time T because the angle between the wave vectors $\vec{k}$ and the rod axis OX varies between zero and π, subject to the restriction $0 \leqslant \sigma \leqslant \sigma_{max}$. The maximum value of the angle σ is simply the angle subtended by the bolometer

$$\sigma_{max} = \tan^{-1}(d/\ell), \tag{2}$$

where d is the diameter of the quartz rod and ℓ is the length of the quartz rod.

The salient effect of the elastic mismatch between the metallic generating and receiving films and the quartz propagation medium manifests itself in the time constants of the generator and the detector. At low temperatures a thermal discontinuity, ΔT, exists at the interface between any two solid media in thermal contact whenever a heat flux passes across the interface. This so-called thermal boundary resistance, R_B^*, is similar to the Kapitza resistance between liquid helium and a solid and is defined as

$$\dot{Q} = \frac{A\Delta T}{R_B^*}, \tag{3}$$

where $\dot{Q}$ is the rate of heat flow (watts), A is the area of interface (cm^2), ΔT is the thermal discontinuity ($°K$), and R_B^* is the thermal boundary resistance ($°K\,cm^2\,watts^{-1}$). A theoretical expression for the thermal boundary resistance has been derived by W. A. Little,[5] and is expressed as a function of the elastic properties of the adjoining media. Experimental values for the boundary resistance between indium and sapphire have been obtained by Neeper and Dillinger.[6] The results of their experiments compare favorably with an extrapolation of the theory of Little; hence, we have estimated the boundary resistance of a quartz-tin interface, using a similar extrapolation. Thus

$$R_B^* = 8T^{-3} \; °K \, cm^2 \, watts^{-1} \qquad \text{(estimated for quartz-tin).} \tag{4}$$

It is not entirely obvious that the Kapitza resistance between a liquid helium-metal interface should have the same order of magnitude as the thermal boundary resistance between quartz and tin; yet with gold and copper this has been the experimental observation[7]

$$R_K^* = 8T^{-3} \; °K \, cm^2 \, watts^{-1} \qquad \text{(observed for gold-liquid helium).} \tag{5}$$

We have no data on the Kapitza resistance of tin; we assume that the value for gold represents a very good approximation. The thermal time constant of a metallic film on a solid substrate submerged in liquid helium is determined by the parallel combination of the thermal boundary resistance and the Kapitza resistance.

$$\tau = \rho c_v d \, \frac{R_K^* R_B^*}{R_K^* + R_B^*} \; \text{sec},\tag{6}$$

where ρ is the density of the metallic film (gm cm^{-3}), c_v is the heat capacity of the metallic film (J gm^{-1} °K^{-1}), and d is the thickness of the metallic film (cm). Using recent data on the low-temperature heat capacity of tin, we find that the time constant of a 1000 Å film at 3.6°K is 10^{-9} sec. This figure supports our previous contention that the pulse broadening is not associated with the thermal response time of the bolometer.

In a subsequent experiment, a layer of glue, ~10^{-3} cm thick, was painted over the bolometer. Typical heat capacities of such materials are ten times greater than those for metals at helium temperature. We have no data on the thermal boundary and Kapitza resistances for glue, but perhaps it is relevant to point out that experimental values of these resistances vary less than an order of magnitude over a wide range of materials.[7] Assuming that the thermal resistances associated with the glue are not significantly different, we obtain a time constant, $\tau \sim 10^{-6}$ sec, for the glue-coated bolometer. Figure II-4 shows the pulses of incoherent phonons intercepted by the glue-coated bolometer. As expected, most of the mode structure has been lost.
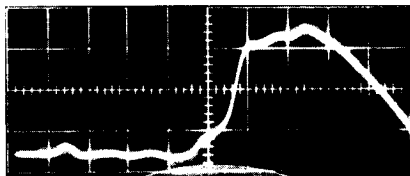
Fig. II-4.   Incoherent phonons in x-cut quartz observed with a glue-coated bolometer.

Using an expression for the thermal discontinuity across an interface that is valid when the temperature difference is large compared with the ambient temperature,[5] we can estimate the temperature rise of our generating film.

$$\dot{Q} = \frac{A\left(T^4 - T_o^4\right)}{4R^* T^3}.\tag{7}$$

From VSWR measurements on the aluminum-generating film, we have determined that approximately 20 per cent of the input power is converted into heat in the film. Therefore, for a total power input of $\dot{Q} \approx 40$ watts, we find $T \sim 7$°K. (This is sufficiently high to call into question our tacit assumption of thermal equilibrium.) If we also assume that the frequency spectrum of the thermal vibration in our generating film can be characterized by a Planck distribution, we can estimate the order of magnitude of the dominant frequencies involved in the transport of thermal energy through the quartz rod:

$$\nu_{peak} \approx \frac{3kT}{h}. \tag{8}$$

The peak of the Planck distribution occurs at $\nu_{peak} \approx 0.4 \times 10^{12}$ sec^{-1} in our experiment. The lattice constant of quartz is approximately 5 Å. This places the edge of the first Brillouin zone at 6.3×10^{7} cm^{-1}. Thus, the maximum frequency that can be propagated on the longitudinal branch of the acoustic mode is given by

$$\nu_{max} = 5.7 \times 10^{12} \text{ sec}^{-1}. \tag{9}$$

For the slow-transverse branch,

$$\nu_{max} = 3.3 \times 10^{12} \text{ sec}^{-1}. \tag{10}$$

These considerations serve to point out a second possible origin of the signal-pulse broadening that is evident in Fig. II-1. From the foregoing analysis, it appears that a significant portion of the elastic energy of these incoherent phonon pulses is carried by modes whose frequencies are close to the cutoff frequency of the lattice. Whether dispersion effects are making an appreciable contribution to the pulse broadening is a subject for further investigation.

At low temperatures the mean-free path of a phonon in a crystal approaches the dimension of the crystal. Our observations are consistent with rectilinear propagation of the phonons, without evidence of the dominant scattering mechanism that characterizes heat flow in solids at room temperature. Thus we have observed the breakdown of the diffusion equation. As we have mentioned, approximately 20 per cent of our total input power is dissipated in the aluminum generating film. Of this, half flows directly to the liquid helium through the Kapitza resistance. Thus, for a peak input power of 40 watts, only ~4 watts enters the quartz rod through the thermal boundary resistance. Because of rectilinear propagation, only 4×10^{-3} watt intercepts the receiving bolometer directly, if we assume that the thermal energy at the generating film is distributed evenly over a solid angle of 2π.

In order to estimate the resulting voltage rise in the bolometer, we draw upon an expression for the responsivity of a bolometer.[8]

$$r = \frac{\Delta V}{\Delta \dot{Q}} = \frac{i\alpha R R^*}{2A}, \tag{11}$$

where ΔV is the voltage rise (volts), $\Delta \dot{Q}$ is the power input (watts), i is the bias current (amps), $\alpha = \frac{1}{R} \frac{dR}{dT}$ is the temperature coefficient of resistivity (°K^{-1}), R is the resistance of the bolometer (ohms), R^* is the effective thermal resistance of bolometric film (°K cm^2 watts^{-1}), and A is the area of the bolometric film (cm^2).

In Eq. 11 we have assumed that the bolometer is terminated in a load impedance
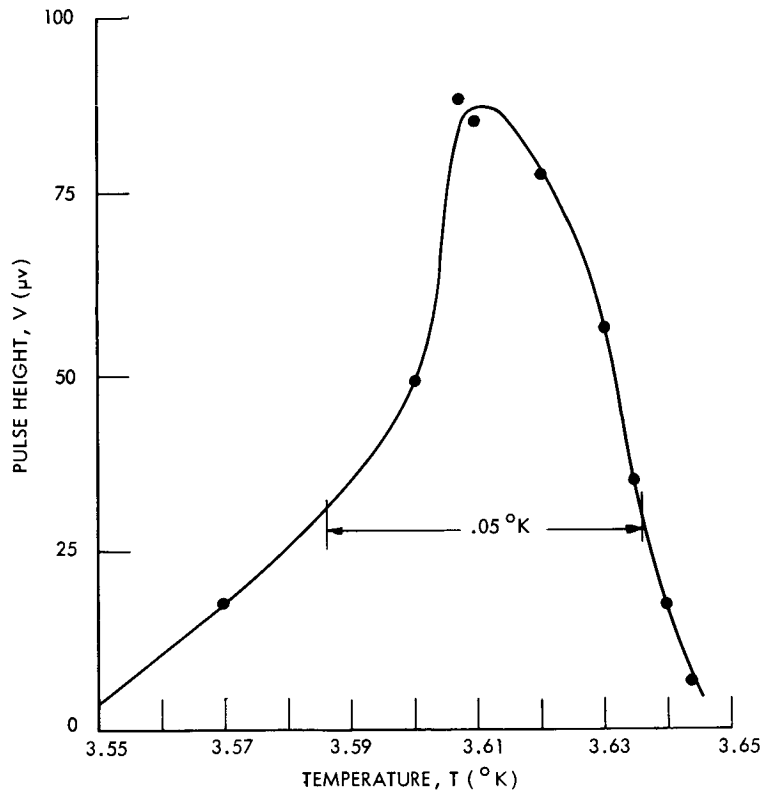
Fig. II-5. Signal voltage as a function of bolometer temperature. The peak occurs
at the maximum slope of the resistivity for the superconducting transition
of the bolometer.

.equal to its own resistance, R. In order to estimate a, we plotted the received signal
as a function of bolometer (ambient) temperature. This is shown in Fig. II-5. We see
that the sensitive region of the bolometer occurs within a temperature range of $0.05°K$.
Since $a \approx (\Delta T)^{-1}$, we estimate $a \approx 20$. The resistance of the bolometer, R, is $\sim 0.1\,\Omega$,
and the bias current was 38 ma for our observations. The effective thermal boundary
resistance, previously calculated, is $\sim 4T^{-3}$, and the area, A, of the bolometer is
$0.07\,cm^2$. These values, substituted in Eq. 11 yield a responsivity, r ~ 0.05 volts/watt.
With 4×10^{-3} watt incident upon the bolometer, we expect a voltage rise of the order of
200 µv. Within the approximate nature of our calculation, this compares favorably with
the observed voltage of 90 µv.

The author wishes to express appreciation to Professor J. F. Cochran, who pointed
out the importance of the thermal-boundary resistance and the Kapitza resistance, and
to acknowledge the assistance of Mr. M. C. Graham, who took part in the experi-
ments.

J. M. Andrews, Jr.

1. R. J. von Gutfeld and A. H. Nethercot, Jr., Phys. Rev. Letters 12, 641 (1964).

2. J. M. Andrews, Jr., Incoherent phonon propagation in anisotropic media, Quarterly Progress Report No. 75, Research Laboratory of Electronics, M.I.T., October 15, 1964, pp. 5-7.

3. J. M. Andrews, Jr., Ultrasonic attenuation in superconductors, Quarterly Progress Report No. 70, Research Laboratory of Electronics, M.I.T., July 15, 1963, pp. 8-17.

4. M. J. P. Musgrave, Rept. Progr. Phys. 22, 74 (1959).

5. W. A. Little, Can. J. Phys. 37, 334 (1959).

6. D. A. Neeper and J. R. Dillinger, Phys. Rev. 135, A1028 (1964).

7. R. C. Johnson and W. A. Little, Phys. Rev. 130, 596 (1963).

8. R. A. Smith, F. E. Jones, and R. P. Chasmar, The Detection and Measurement of Infrared Radiation (Oxford University Press, London, 1957), p. 99.

III. RADIO ASTRONOMY[*]

Prof. A. H. Barrett
Prof. J. W. Graham
Prof. M. Loewenthal
Prof. R. P. Rafuse
Dr. D. H. Staelin
Dr. S. H. Zisk

R. J. Allen
R. K. Breon
Patricia P. Crowther
A. B. Hull
W. B. Lenoir
J. M. Moran, Jr.

M. A. Palfy
S. M. Rezende
A. E. E. Rogers
J. H. Spoor
D. H. Steinbrecher
A. Vander Vorst

A. MEASUREMENTS OF THE MICROWAVE SPECTRUM OF VENUS NEAR 1-cm WAVELENGTH

During June and July, 1964, observations were made of the planet Venus[1] at 9-14 mm wavelengths. These observations were made with the use of the Research Laboratory of Electronics five-channel microwave radiometer[2] mounted in the 28-ft millimeter wavelength antenna at Lincoln Laboratory, M.I.T. The frequencies 21.9, 23.5, 25.5, 29.5, and 32.4 Gc/sec were observed simultaneously; the frequency 21.1 Gc/sec was observed separately with a one-channel radiometer.

The spectral measurements were made by comparing the radio spectrum of Venus with that of the moon. The moon was observed on 17 days during the experimental period. In order to relate the observations of Venus to those of the moon, antenna patterns and values of the atmospheric absorption as a function of frequency were required. The antenna pattern at each frequency was measured with a test signal source mounted on a tower, 6 miles from the antenna site. The antenna was suitably defocused for these measurements, and cross-polarization patterns were also measured. The atmospheric opacity was determined by a series of solar extinction measurements which were used to relate the opacity to the ground-level humidity. This empirically determined relationship was then used to determine the appropriate atmospheric opacity at any given time.

The results obtained on each day are summarized in Table III-1. In the table, T_{BV}(°K) is the average brightness temperature of the visible disk of Venus. σ(%) is the estimated standard deviation of the measured signal from its true value and does not include uncertainties introduced by atmospheric absorption, antenna pointing, and so forth.

If all of the separate observations are averaged together, and each is weighted by its estimated accuracy, then the results listed in Table III-2 are obtained.

The absolute error presented in Table III-2 includes all sources of error; measurements made at different frequencies are considered to be independent. The relative error includes only those components of error that are independent from channel to channel.

Table III-1.   Observed brightness temperature of Venus as a function of date.

Date 1964	32.4 (Gc/sec)		29.5		25.5		23.5		21.9		21.1	
	T_{BV}(°K)	σ(%)	T_{BV}	σ	T_{BV}	σ	T_{BV}	σ	T_{BV}	σ	T_{BV}	σ
6/5	458	3	501	10	518	6	445	5	353	17	—	
6/11	418	5	415	12	449	9	421	7	436	14	—	
6/12	403	7	463	8	442	13	440	13	389	21	—	
6/12	420	10	395	14	449	7	482	5	358	24	—	
6/29	513	17	307	17	482	12	412	7	527	25	—	
7/5	467	52	584	8	—		406	7	350	14	—	
7/6	522	17	549	16	439	6	513	5	482	12	—	
7/7	422	13	400	25	493	9	427	9	377	22	—	
7/8	428	6	505	10	403	5	381	10	375	6	—	
7/11	312	18	421	22	344	8	518	6	359	18	—	
7/15	—		—		—		—		—		495	12
7/16	477	8	457	17	409	6	220	30	—		—	
7/17	—		—		—		—		—		524	13
7/18	—		—		—		—		—		528	9
7/27	322	23	499	40	385	8	474	18	518	12	—	
7/30	346	17	374	44	417	10	535	11	491	14	—	

Table III-2. Averaged and weighted results.

(Gc/sec)	T_{BV}(°K)	Relative error (°K)	Absolute error (°K)
32. 4	430	±24	±42
29. 5	463	±32	±68
25. 5	428	±20	±46
23. 5	450	±23	±41
21. 9	404	±28	±39
21. 1	502	±82	±100

The results are also presented in Fig. III-1, together with the results of other observers made during several inferior conjunctions. One theoretical spectrum was computed for a cloud layer that is uniform from 465°K to 270°K, with absorption coefficient proportional to the square of the frequency. A similar spectrum would be expected from water clouds. The second theoretical spectrum is for nonresonant absorption by a 10% CO_2-90% N_2 atmosphere with surface pressure approximately 150 atmospheres.

The 32. 4 and 29. 5 Gc/sec data points are 1. 5 and 1. 7 standard deviations above the nonresonant spectrum, respectively, and the 21. 9 Gc/sec measurement is 2. 0 standard deviations below the same curve. Even if the absolute error brackets are used, the
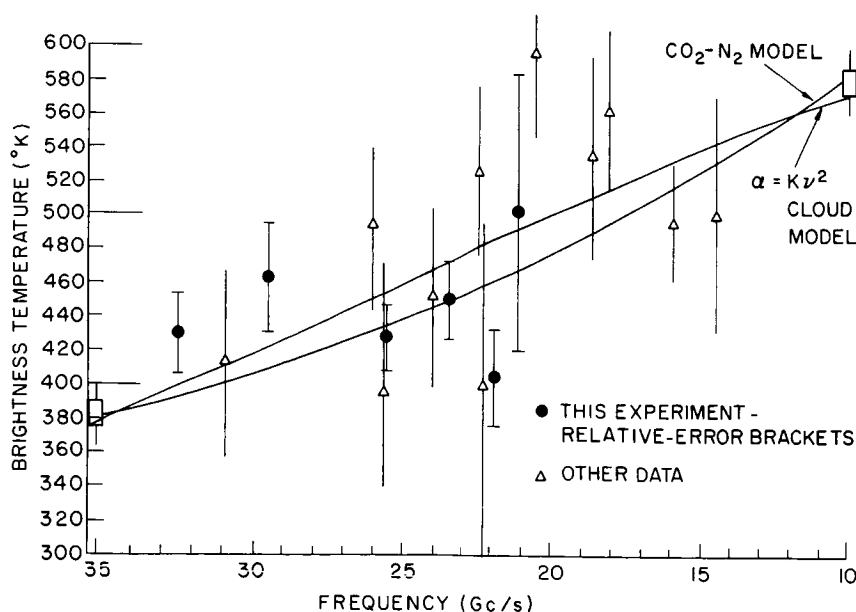
Fig. III-1. Microwave spectrum of Venus near 1-cm wavelength.

displacements are 0. 9, 0. 8, and 1. 4σ, respectively. The probability of being farther than 1. 5 standard deviations from the correct value is approximately 0. 13, and the prob-ability of being farther than 2. 0 standard deviations is approximately 0. 05. Thus if the relative error brackets are used, it is unlikely that the microwave spectrum of Venus is nonresonant in character over this spectral region.

Atmospheric models in better agreement with the data are those having molecular resonances in the millimeter wavelength region. Also, models incorporating scattering are in agreement. A study of these models is in progress.

D. H. Staelin, A. H. Barrett

References

1. D. H. Staelin, Sc. D. Thesis, Department of Electrical Engineering, Massachu-setts Institute of Technology, January 1965.

2. D. H. Staelin, Quarterly Progress Report No. 69, Research Laboratory of Elec-tronics, M. I. T., April 15, 1963, pp. 23-25.

B. OBSERVATIONS OF MICROWAVE EMISSION FROM ATMOSPHERIC OXYGEN

Two more balloon flights[1,2] were undertaken from Palestine, Texas, in the fall of 1964. The first, Flight No. 88P, on 29 October, was partially successful. The second, Flight No. 89P, on 8 November, was completely successful, yielding good data for the entire flight.

1. Flight Radiometer

The radiometer is basically the same as the one used in previous flights. See Fig. III-2. For these flights the telemeter was operating, giving a real-time data output.

The errors are the sum of a consistent error (constant over the duration of the flight) and a random noise. The consistent error varies from ±8°K for a brightness tem-perature of 0°K to ±1°K for a brightness temperature of 300°K. The rms temperature variation of the radiometer is approximately 1°K on each channel.

2. Results

Flight No. 88P, 29 October 1964

This flight was only partially successful. Data were recovered from the 20-Mc and 200-Mc IF channels during a period of 2 hours (approximately the time for ascending to float altitude 30 km), and from the 60-Mc IF channel for the first 20 minutes and the last 20 minutes of the first 2 hours.

After 2 hours, the programmer assumed a supposedly unallowed state and began
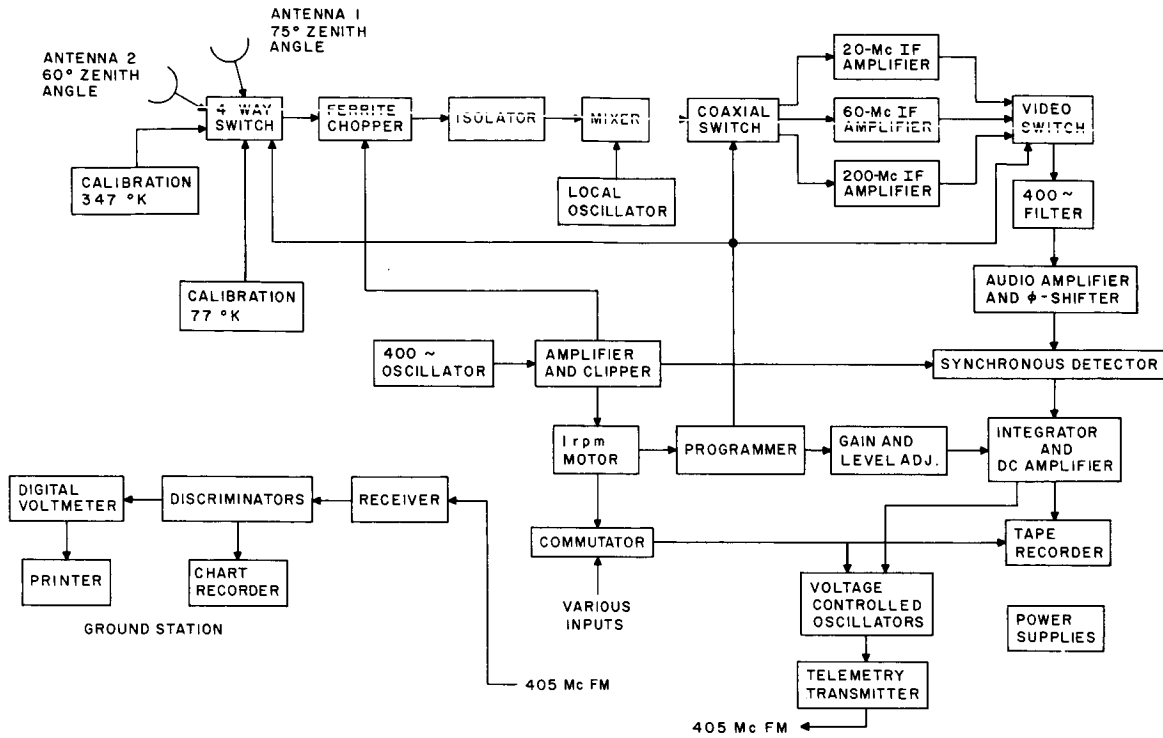
Fig. III-2. Flight radiometer.

switching between the calibration loads only. The difficulty in the 60-Mc IF channel was traced to a faulty relay which evidently failed after 20 minutes and then began to perform properly after one hour of failure.

Telemetry apparatus permitted the malfunctions to be viewed as they were happening. As soon as the programmer malfunction was diagnosed, the flight was terminated.

The small amount of data from flight No. 88P has not been reduced yet.
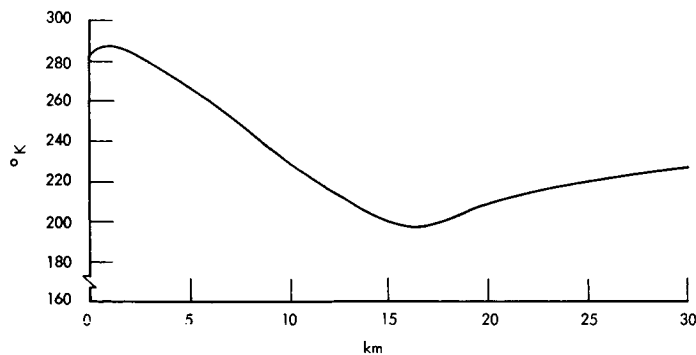
Fig. III-3. Flight No. 89P atmospheric temperature vs height.

Flight No. 89P, 8 November 1964

The erratic relay was replaced and the programmer's difficulty was solved by use of a mechanical commutator. The profile of the flight was the following.

1. Approximate linear ascent to 30 km in 2 hours.
2. Float at 30 km for 4 hours.
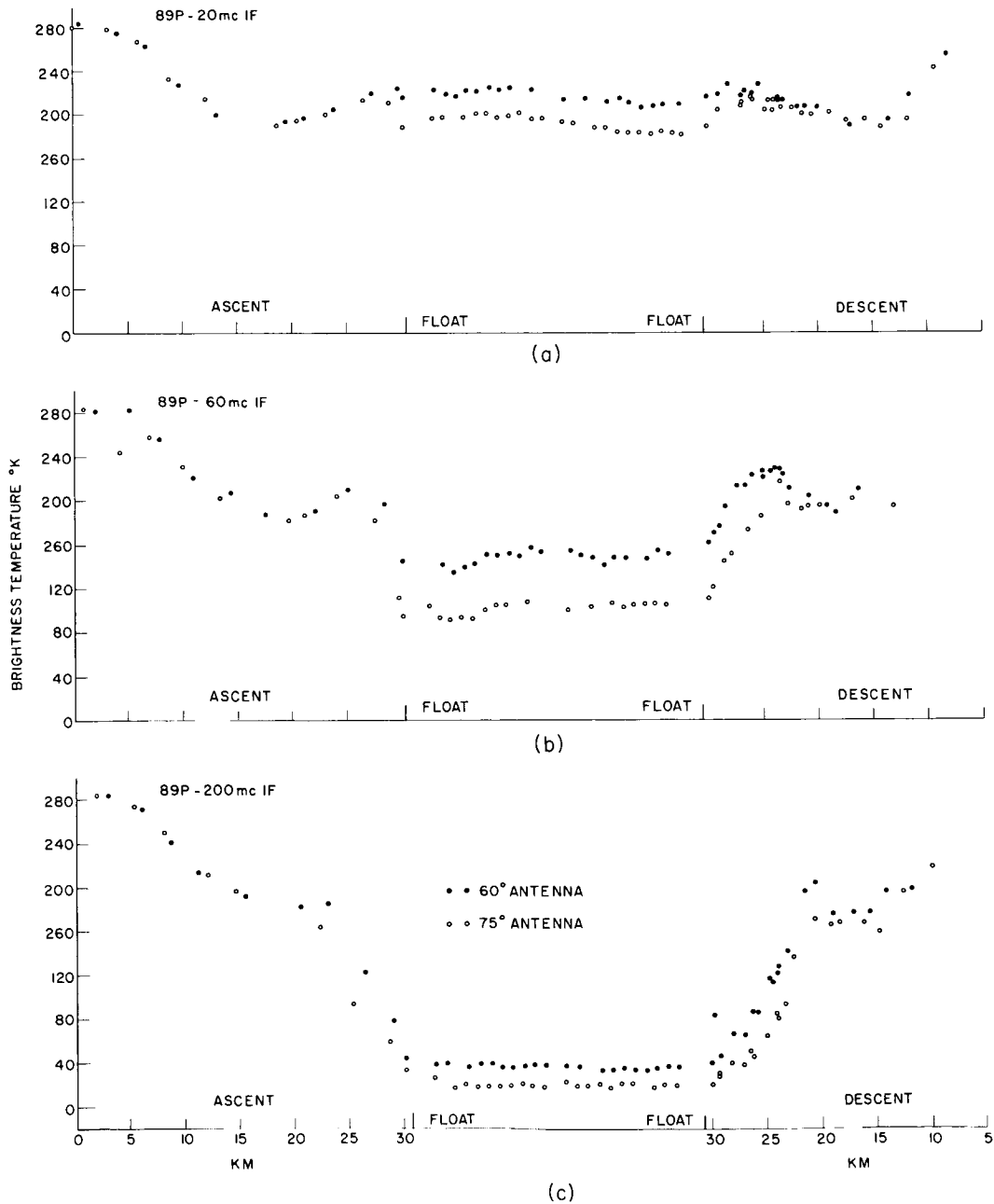
Fig. III-4. Flight No. 89P brightness temperatures vs height.

3. Approximate linear descent to 5.5 km in 3 hours.

4. Parachute to ground.

The atmospheric temperature versus height as measured during the flight is shown in Fig. III-3.

Data were taken in three ways:

1. At the balloon base with the master receiver as indicated in the block diagram of Fig. III-2.

2. In a car, equipped with a receiver and a chart recorder, following the path of the balloon.

3. On a tape recorder in the flight gondola.

Data taken by method 1 were the most accurate because of the resolution it afforded. Data taken by method 2 was the least accurate because the chart recording could only be read to ±5°K, while the radiometer noise was only ±1°K. Data taken by method 2 should eventually be as accurate as that taken by method 1. Tape-recorder malfunctions have prevented getting all of the data from the tape.

The brightness temperatures versus height for the three IF channels and the two antenna angles are shown in Fig. III-4.

The six brightness temperatures averaged over the duration of the flight at float altitude are given in Table III-3.

Table III-3.   Float brightness temperature averages.

Channel	60° Antenna	75° Antenna
20 Mc IF	190°K	215°K
60 Mc IF	102°K	148°K
200 Mc IF	19°K	36°K

Work continues on further interpretation of these data. In particular, work is under way to see how much information can be obtained about the atmospheric thermal structure above float height, to see what can be inferred about the linewidth, and what can be said about the line intensity.

Another series of flights is planned tentatively for Summer 1965 with improvements that should lower the temperature sensitivity to less than 2°K for all brightness temperatures.

W. B. Lenoir

References

1. W. B. Lenoir and J. W. Kuiper, Quarterly Progress Report No. 75, Research Laboratory of Electronics, M.I.T., October 15, 1964, pp. 9-18.

2. A. H. Barrett, J. C. Blinn III, and J. W. Kuiper, Quarterly Progress Report No. 71, Research Laboratory of Electronics, M.I.T., October 15, 1963, pp. 69-76.

C. MATRIX FORMULATION OF RADIATIVE TRANSFER

In many cases of practical interest the emission and absorption properties of a medium depend on the polarization of the radiation. In these cases, for a general treatment, it does not suffice to treat the radiative transfer in the framework of the scalar equation of radiative transfer. A new treatment must be developed which includes polarization information, as well as intensity information. This report concerns such a development. It is assumed that a spatially and angularly incoherent TEM wave traveling in the $+\underline{Z}$ direction is being dealt with.

1. Coherency Spectrum Matrix

The electric field can be written

$$\vec{\mathscr{E}}(t) = \vec{\mathscr{E}}_\alpha(t) + \vec{\mathscr{E}}_\beta(t) \tag{1}$$

in which the subscripts α and β indicate the components of $\vec{\mathscr{E}}(t)$ with polarization α and β, respectively, with α and β being any two opposite polarizations. They will be called a "polarization basis."

Samples of $\mathscr{E}_\alpha(t)$ and $\mathscr{E}_\beta(t)$ that are of T duration will have Fourier Transforms given by

$$E_{\alpha,T}(\nu) = \int_{-T/2}^{T/2} \mathscr{E}_\alpha(t)\, e^{-i2\pi\nu t}\, dt \tag{2a}$$

$$E_{\beta,T}(\nu) = \int_{-T/2}^{T/2} \mathscr{E}_\beta(t)\, e^{-i2\pi\nu t}\, dt. \tag{2b}$$

The Fourier transform of the corresponding sample of $\vec{E}_T(t)$, also of T duration, can be written as a vector in the two-dimensional vector space of polarizations.

$$E_T(\nu) = \begin{pmatrix} E_{\alpha,T}(\nu) \\ E_{\beta,T}(\nu) \end{pmatrix} \tag{3}$$

A "coherency spectrum matrix" can now be defined as

$$\underline{\underline{J}}(\nu) = \left(\lim_{T\to\infty} \overline{\frac{E_T(\nu)\, E_T^{t*}(\nu)}{T}} \right) \tag{4}$$

in which $\underline{\underline{J}}$ denotes that J is a matrix, t denotes transpose, * denotes complex conjugate, and the bar over the relation denotes the ensemble average.[1] This $\underline{\underline{J}}(\nu)$ is related to the coherency matrix of Wolf[2] but does not require a narrow-band assumption.

Substitution of (3) in (4) yields

$$\underline{J}(\nu) = \begin{pmatrix} \left[\lim_{T \to \infty} \dfrac{|E_{\alpha,T}(\nu)|^2}{T} \right] & \left[\lim_{T \to \infty} \dfrac{E_{\alpha,T}(\nu) E^*_{\beta,T}(\nu)}{T} \right] \\ \left[\lim_{T \to \infty} \dfrac{E^*_{\alpha,T}(\nu) E_{\beta,T}(\nu)}{T} \right] & \left[\lim_{T \to \infty} \dfrac{|E_{\beta,T}(\nu)|^2}{T} \right] \end{pmatrix}. \tag{5}$$

Through use of the relation for power spectral density

$$\Phi_{ij}(\nu) = \left[\lim_{T \to \infty} \frac{E_{i,T}(\nu) E^*_{j,T}(\nu)}{T} \right], \tag{6}$$

Eq. 5 becomes

$$\underline{J}(\nu) = \begin{pmatrix} \Phi_{\alpha\alpha}(\nu) & \Phi_{\alpha\beta}(\nu) \\ \Phi_{\beta\alpha}(\nu) & \Phi_{\beta\beta}(\nu) \end{pmatrix} \tag{7}$$

which is more general, since $\Phi_{ij}(\nu)$ is defined for random noiselike fields, whereas $E(\nu)$ is not.

Examination of (7), or equivalently of (5), shows at once that $\underline{J}(\nu)$ is Hermitian (self-adjoint), that is, $\underline{J}(\nu) = \underline{J}(\nu)^{t*}$. It is also obvious that $\Phi_{\alpha\alpha}(\nu)$ and $\Phi_{\beta\beta}(\nu)$ are real and non-negative. $\Phi_{\alpha\alpha}(\nu)$ and $\Phi_{\beta\beta}(\nu)$ are the power spectral densities in polarizations α and β, respectively. Hence the trace of $\underline{J}(\nu)$,

$$\text{tr } \underline{J}(\nu) = \Phi_{\alpha\alpha}(\nu) + \Phi_{\beta\beta}(\nu) \geq 0, \tag{8}$$

is the total power spectral density of the radiation.

The off-diagonal terms, $\Phi_{\alpha\beta}(\nu)$ and $\Phi_{\beta\alpha}(\nu)$, measure the degree of coherence between the radiation with polarization α and that with polarization β. By Schwartz' inequality, the determinant of $\underline{J}(\nu)$,

$$\det \underline{J}(\nu) = \Phi_{\alpha\alpha}(\nu) \Phi_{\beta\beta}(\nu) - \Phi_{\alpha\beta}(\nu) \Phi_{\beta\alpha}(\nu), \tag{9}$$

is real and non-negative.

The analysis thus far has assumed α,β to be the polarization basis. The change from one polarization basis α,β to another, x,y, is effected through a unitary transformation, U.

$$\begin{pmatrix} E_{x,T}(\nu) \\ E_{y,T}(\nu) \end{pmatrix} = \underline{U} \begin{pmatrix} E_{\alpha,T}(\nu) \\ E_{\beta,T}(\nu) \end{pmatrix} \tag{10}$$

This changes only the polarization basis in which the radiation is described. It does not change the radiation itself in any way.

From (5) and (10), the coherency spectrum matrix is transformed to the basis x, y

through

$$\underline{\underline{J}}_{x,y}(\nu) = \underline{\underline{U}}\, \underline{\underline{J}}_{\alpha,\beta}(\nu)\, \underline{\underline{U}}^{t*}. \tag{11}$$

From which it is obvious that $\underline{\underline{J}}_{x,y}(\nu)$ is Hermitian, since $\underline{\underline{J}}_{\alpha,\beta}(\nu)$ is.

From the properties of unitary transformations it can be shown that the trace, determinant, and eigenvalues are invariant under a transformation such as (11). So that $\operatorname{tr} \underline{\underline{J}}(\nu)$, $\det \underline{\underline{J}}(\nu)$, $\lambda_1(\nu)$, $\lambda_2(\nu)$ are the trace, determinant, and eigenvalues of $\underline{\underline{J}}(\nu)$ in any polarization basis.

Then there exists a unitary transformation, $U_D(\nu)$ to a polarization basis, m, n in which $\underline{\underline{J}}(\nu)$ is diagonal.

$$\underline{\underline{J}}_{m,n}(\nu) = \underline{\underline{U}}_D(\nu)\, \underline{\underline{J}}_{\alpha,\beta}(\nu)\, \underline{\underline{U}}_D^{t*}(\nu) = \begin{pmatrix} J_m(\nu) & 0 \\ 0 & J_n(\nu) \end{pmatrix} \tag{12}$$

with $J_m(\nu) = \lambda_1(\nu)$ and $J_n(\nu) = \lambda_2(\nu)$. Assuming, with no loss of generality, that $J_m(\nu) \geqslant J_n(\nu)$, we have

$$\underline{\underline{J}}_{m,n}(\nu) = J_n(\nu) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + [J_m(\nu) - J_n(\nu)] \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{13}$$

where both matrices on the right are valid coherency spectrum matrices.

The first is seen to be that of a randomly polarized (unpolarized) wave; while the second is that of a totally polarized wave with polarization m. Such a decomposition exists uniquely (through a function of ν in general) for all coherency spectrum matrices.

The fractional polarization spectrum can now be defined as the power spectral density of the polarized part divided by the total power spectral density

$$p(\nu) = \frac{J_m(\nu) - J_n(\nu)}{J_m(\nu) + J_n(\nu)}. \tag{14}$$

Here, $p(\nu)$ is independent of the polarization basis and is given more generally by

$$p(\nu) = \sqrt{1 - \frac{4 \det \underline{\underline{J}}(\nu)}{[\operatorname{tr} \underline{\underline{J}}(\nu)]^2}}. \tag{15}$$

In radio astronomy it is often convenient to use the concept of brightness temperature, rather than intensity or spectral flux density. For the scalar description, the brightness temperature (in the microwave region with $h\nu \ll kT$) is defined by

$$I(\nu) = \frac{2k\nu^2 T_B(\nu)}{c^2}, \tag{16}$$

where I is the intensity; T_B, the brightness temperature; k, Boltzmann's constant; and c, the velocity of light.

It will also be convenient to define a brightness-temperature coherency spectrum matrix in similar fashion. Since $2k\nu^2/c^2$ is constant (for a given ν), and $\underline{J}(\nu)$ is the intensity coherency spectrum matrix (except for a multiplicative constant), we may define

$$\underline{\underline{T}}_B(\nu) = \text{const}(\nu)\,\underline{\underline{J}}(\nu) \tag{17}$$

as the brightness-temperature coherency spectrum matrix. Thus on the polarization basis, a, β

$$\underline{\underline{T}}_B(\nu) = \begin{pmatrix} T_{aa}(\nu) & T_{a\beta_R}(\nu) + iT_{a\beta_I}(\nu) \\ T_{a\beta_R}(\nu) - iT_{a\beta_I}(\nu) & T_{\beta\beta}(\nu) \end{pmatrix}, \tag{18}$$

where $T_{aa}(\nu)$ and $T_{\beta\beta}(\nu)$ are the brightness temperatures (in a scalar sense) of the radiation with polarizations a and β. The general properties of $\underline{J}(\nu)$ apply also to $\underline{\underline{T}}_B(\nu)$.

In particular, the fraction polarization spectrum, $p(\nu)$, is unchanged.

$$p(\nu) = \sqrt{1 - \frac{4\det \underline{\underline{T}}_B(\nu)}{[\operatorname{tr} \underline{\underline{T}}_B(\nu)]^2}}\,. \tag{19}$$

2. Matrix Equation of Radiative Transfer

The propagation of the brightness-temperature coherency spectrum matrix through a medium will now be considered. The medium is assumed to be only slightly inhomogeneous (the properties of the medium vary only slightly over a wavelength's distance).

Then a complex propagation matrix, $\underline{\underline{A}}(\nu)$, that is the matrix equivalent of the scalar
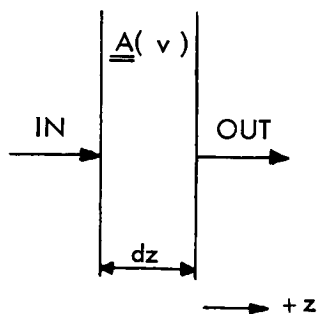
Fig. III-5. Geometry for a slab of infinitesimal thickness.

complex propagation constant can be defined. $\underline{\underline{A}}(\nu)$ describes the absorption and propagation properties of the medium and, in general, will be a function of position (subject

to the slightly inhomogeneous assumption).

The relations for the geometry of Fig. III-5 are

$$E_{out}(\nu) = \left[\underline{\underline{I}} - \underline{\underline{A}}(\nu) \, dz\right] E_{in}(\nu) \tag{20}$$

with $\underline{\underline{I}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\underline{\underline{A}}(\nu)$ is the complex propagation matrix. This equation is consistent with Maxwell's equations (assuming small loss and only slight inhomogeneities). $\underline{\underline{A}}(\nu)$ can be found in terms of $\underline{\underline{\epsilon}}(\nu)$, $\underline{\underline{\mu}}(\nu)$, $\underline{\underline{\sigma}}(\nu)$ which appear in the matrix formulation of Maxwell's equations.

From (20) and (5) we obtain

$$\underline{\underline{J}}_{out}(\nu) = \underline{\underline{J}}_{in}(\nu) - \underline{\underline{A}}(\nu) \, \underline{\underline{J}}_{in}(\nu) \, dz - \underline{\underline{J}}_{in}(\nu) \, \underline{\underline{A}}^{t*}(\nu) \, dz + \underline{\underline{A}}(\nu) \underline{\underline{J}}_{in}(\nu) \, \underline{\underline{A}}^{t*}(\nu) \, dz^2.$$

Letting $d\underline{\underline{J}}(\nu) = \underline{\underline{J}}_{out}(\nu) - \underline{\underline{J}}_{in}(\nu)$ and taking $\lim_{dz \to 0}$ yields

$$\frac{d}{dz} \underline{\underline{J}}(\nu) + \underline{\underline{A}}(\nu) \, \underline{\underline{J}}(\nu) + \underline{\underline{J}}(\nu) \, \underline{\underline{A}}^{t*}(\nu) = 0. \tag{21}$$

Equation 21 is the matrix equation of radiative transfer in which emission from the medium has been ignored.

The equivalent equation in brightness temperature notation is

$$\frac{d}{dz} \underline{\underline{T}}_B(\nu) + \underline{\underline{A}}(\nu) \, \underline{\underline{T}}_B(\nu) + \underline{\underline{T}}_B(\nu) \, \underline{\underline{A}}^{t*}(\nu) = 0. \tag{22}$$

Consider the propagation through a slab of finite thickness with $\underline{\underline{A}}(\nu)$ independent of position (see Fig. III-6). The problem is to solve (22) for an incident brightness temperature coherency spectrum matrix of $\underline{\underline{T}}_{B_i}(\nu)$.
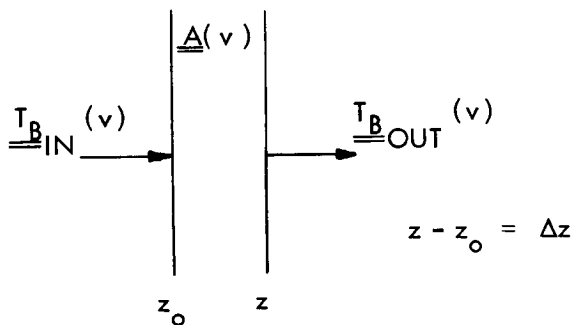
Fig. III-6. Geometry for a slab of finite thickness.

$\underline{\underline{T}}_{B_i}(\nu)$ is not sufficient to specify $E_i(\nu)$; nevertheless it is often useful to consider $E_i(\nu)$ as if it were known and to convert to $\underline{\underline{T}}_{B_o}(\nu)$ only at the last step. As long as claims are made only to $\underline{\underline{T}}_{B_o}(\nu)$ and not to $E_o(\nu)$, this is a valid procedure.

In this manner Eq. 20 is readily solved to yield

$$E_0(\nu) = e^{-\underline{\underline{A}}(\nu)\Delta z} E_i(\nu),$$ (23)

where the exponential of a matrix is defined by the power series

$$e^{-\underline{\underline{A}}(\nu)\Delta z} = \sum_{n=0}^{\infty} \frac{[-\underline{\underline{A}}(\nu)\Delta z]^n}{n!}; \quad \text{with } \underline{\underline{A}}(\nu)^0 = \underline{\underline{I}}.$$ (24)

A word of warning is in order here. Considerable care must be exercised when dealing with exponentials of matrices. Many of the familiar relations and rules governing exponentials of scalars do not apply to the exponentials of matrices. For example, if $\underline{\underline{X}}$ and $\underline{\underline{Y}}$ are matrices, then $e^{\underline{\underline{X}}+\underline{\underline{Y}}} = e^{\underline{\underline{X}}} e^{\underline{\underline{Y}}}$ if and only if $\underline{\underline{XY}} = \underline{\underline{YX}}$, that is, $\underline{\underline{X}}$ and $\underline{\underline{Y}}$ commute.

Using (23), we obtain the solution for $\underline{\underline{T}}_{B_0}(\nu)$:

$$\underline{\underline{T}}_{B_0}(\nu) = e^{-\underline{\underline{A}}(\nu)\Delta z} \underline{\underline{T}}_{B_i}(\nu) e^{-\underline{\underline{A}}^{t*}(\nu)\Delta z}.$$ (25)

Thus far the possibility of emission from the medium has been ignored. To allow for it Eq. 22 can be rewritten as

$$\frac{d}{dz}\underline{\underline{T}}_B(\nu) + \underline{\underline{A}}(\nu) \underline{\underline{T}}_B(\nu) + \underline{\underline{T}}_B(\nu) \underline{\underline{A}}^{t*}(\nu) = \underline{\underline{S}}_e(\nu)$$ (26)

with $\underline{\underline{S}}_e(\nu)$ being the Hermitian emission spectrum matrix.

The problem of finding $\underline{\underline{S}}_e(\nu)$ can be solved through definition of an emission temperature spectrum matrix, $\underline{\underline{T}}_e(\nu)$. Consider a slab as illustrated in Fig. III-6. Then $\underline{\underline{T}}_e(\nu)$ is defined to be the brightness temperature coherency spectrum matrix necessary to fulfill the condition that if $\underline{\underline{T}}_{B_i}(\nu) = \underline{\underline{T}}_e(\nu)$, then $\underline{\underline{T}}_{B_0}(\nu) = \underline{\underline{T}}_e(\nu)$ also, independent of z. This gives

$$\underline{\underline{S}}_e(\nu) = \underline{\underline{A}}(\nu) \underline{\underline{T}}_e(\nu) + \underline{\underline{T}}_e(\nu) \underline{\underline{A}}^{t*}(\nu),$$ (27)

so that the emission spectrum matrix depends on the emission temperature spectrum matrix and the complex propagation properties of the medium. This is the matrix equivalent of the scalar equation[4]

$$j = \gamma T_e$$

in which j is the emission coefficient; γ, the power absorption coefficient; and T_e, the scalar emission temperature.

If the medium is the local thermodynamic equilibrium, then

$$\underline{\underline{T}}_e(\nu) = \underline{\underline{T}}_k = t_k \underline{\underline{I}},$$ (28)

where $\underline{\underline{T}}_k$ is the kinetic temperature matrix, t_k the scalar kinetic temperature, and $\underline{\underline{I}}$ the 2×2 unit matrix.

In general Eq. 26 can be written as

$$\frac{d}{dz} \underline{\underline{T}}_B(\nu) + \underline{\underline{A}}(\nu)\, \underline{\underline{T}}_B(\nu) + \underline{\underline{T}}_B(\nu)\, \underline{\underline{A}}^{t*}(\nu) = \underline{\underline{A}}(\nu)\, \underline{\underline{T}}_e(\nu) + \underline{\underline{T}}_e(\nu)\, \underline{\underline{A}}^{t*}(\nu). \tag{29}$$

When Eq. 28 is valid this becomes

$$\frac{d}{dz} \underline{\underline{T}}_B(\nu) + \underline{\underline{A}}(\nu)\, \underline{\underline{T}}_B(\nu) + \underline{\underline{T}}_B(\nu)\, \underline{\underline{A}}^{t*}(\nu) = t_k[\underline{\underline{A}}(\nu) + \underline{\underline{A}}^{t*}(\nu)]. \tag{30}$$

The solution to Eq. 29 appropriate to Fig. III-6 is

$$\underline{\underline{T}}_{B_o}(\nu) = e^{-\underline{\underline{A}}(\nu)\Delta z}\, \underline{\underline{T}}_{B_i}(\nu)\, e^{-\underline{\underline{A}}^{t*}(\nu)\Delta z} + \left[\underline{\underline{T}}_e(\nu) - e^{-\underline{\underline{A}}(\nu)\Delta z}\, \underline{\underline{T}}_e(\nu)\, e^{-\underline{\underline{A}}^{t*}(\nu)\Delta z} \right]. \tag{31}$$

The first term in Eq. 31 represents the part of $\underline{\underline{T}}_{B_o}(\nu)$ which is due to the $\underline{\underline{T}}_{B_i}(\nu)$ that is incident on the slab. The second term refers to the emission in the interval (z_o, z) and its subsequent propagation through the rest of the slab.

For cases in which $\underline{\underline{T}}_e(\nu) = t_k \underline{\underline{I}}$, the solution becomes

$$\underline{\underline{T}}_{B_o}(\nu) = e^{-\underline{\underline{A}}(\nu)\Delta z}\, \underline{\underline{T}}_{B_i}(\nu)\, e^{-\underline{\underline{A}}^{t*}(\nu)\Delta z} + t_k \left[\underline{\underline{I}} - e^{-\underline{\underline{A}}(\nu)\Delta z}\, e^{-\underline{\underline{A}}^{t*}(\nu)\Delta z} \right]. \tag{32}$$

3. Finite Bandwidth Considerations

Equation 31 integrated over a finite frequency band will yield the brightness temperature coherency matrix for that center frequency, ν_c, and that bandwidth, $\Delta\nu$.

$$\underline{\underline{T}}_{B_o}(\nu_c, \Delta\nu) = \int_{\nu_c - \Delta\nu/2}^{\nu_c + \Delta\nu/2} \underline{\underline{T}}_{B_o}(\nu)\, d\nu. \tag{33}$$

This $\underline{\underline{T}}_{B_o}(\nu_c, \Delta\nu)$ would describe the radiation appropriate to Eq. 31 after it had been passed through an appropriate bandpass filter.

The fractional polarization of $\underline{\underline{T}}_{B_o}(\nu_c, \Delta\nu)$ is given by

$$p(\nu_c, \Delta\nu) = \sqrt{1 - \frac{4 \det \underline{\underline{T}}_{B_o}(\nu_c, \Delta\nu)}{\left[\operatorname{tr} \underline{\underline{T}}_{B_o}(\nu_c, \Delta\nu) \right]^2}}. \tag{34}$$

Note that it is not equal to $\int p(\nu)\, d\nu$.

In general (33) represents quite a formidable integration. Special cases exist in which Eq. 33 can be greatly simplified. Physically, these cases are of great importance as they encompass the narrow-band cases.

If $\underline{A}(\nu)$ exhibits essentially no ν dependence over the interval $\left(\nu_c - \frac{\Delta\nu}{2}, \nu_c + \frac{\Delta\nu}{2}\right)$, then (33) assumes the same form as (31) with all temperature coherency spectrum matrices replaced by their integral over the bandwidth in question. That is,

$$\underline{\underline{T}}_B(\nu) \rightarrow \int_{\nu_c - \Delta\nu/c}^{\nu_c + \Delta\nu/c} \underline{\underline{T}}_B(\nu)\, d\nu. \tag{35}$$

For cases in which $\underline{\underline{T}}_{B_i}(\nu)$ and $\underline{\underline{T}}_e(\nu)$ are essentially independent of temperature over the bandwidth this reduces to

$$\underline{\underline{T}}_B(\nu) \rightarrow \underline{\underline{T}}_B(\nu_c)\, \Delta\nu. \tag{36}$$

4. Complex Propagation Matrix for the Small-Loss Case

The assumption of small loss means that the loss mechanism absorbs an amount of power over a wavelength that is small compared with the total power of the wave. A TEM solution to Maxwell's equations is sought.

$\vec{E}(\vec{r}, \nu)$ and $\vec{H}(\vec{r}, \nu)$ are the Fourier transforms of the electric and magnetic fields at a position indicated by $\vec{r}$. They can be broken into a sum of two components with opposite polarizations.

$$\vec{E}(\vec{r}, \nu) = \vec{E}_1(\vec{r}, \nu) + \vec{E}_2(\vec{r}, \nu) \tag{37a}$$

and

$$\vec{H}(\vec{r}, \nu) = \vec{H}_1(\vec{r}, \nu) + \vec{H}_2(\vec{r}, \nu). \tag{37b}$$

(Note that the subscripts refer to a polarization, not necessarily of a particular spatial direction. For example, if subscript 1 is to denote linear polarization in the x-direction, then $\vec{E}_1(\vec{r}, \nu) = E_1(\vec{r}, \nu)\, \vec{i}_x$, whereas $\vec{H}_1(\vec{r}, \nu) = H_1(\vec{r}, \nu)\, \vec{i}_y$ for $\vec{E}$ and $\vec{H}$ corresponding to a uniform plane wave. It is also possible for subscript 1 to be right-circular polarization and for subscript 2 to be left-circular polarization. This case points out the pitfalls of coupling the spatial vectors with polarization labels.)

The representation

$$\vec{E}(\vec{r}, \nu) = \begin{pmatrix} \vec{E}_1(\vec{r}, \nu) \\ \vec{E}_2(\vec{r}, \nu) \end{pmatrix} \tag{38a}$$

$$\vec{H}(\vec{r}, \nu) = \begin{pmatrix} \vec{H}_1(\vec{r}, \nu) \\ \vec{H}_2(\vec{r}, \nu) \end{pmatrix} \tag{38b}$$

emphasizes that $\vec{E}$ and $\vec{H}$ are vectors in the two-dimensional vector space of polarizations, as well as two-dimensional (TEM) vectors in space.

Allowing the constituency relations to be polarization dependent (but not spatially dependent) gives

$$\vec{B}(\vec{r}, \nu) = \underline{\underline{\mu}}(\nu)\, \vec{H}(\vec{r}, \nu) \tag{39a}$$

$$\vec{D}(\vec{r}, \nu) = \underline{\underline{\epsilon}}(\nu)\, \vec{E}(\vec{r}, \nu) \tag{39b}$$

$$\vec{J}(\vec{r}, \nu) = \underline{\underline{\sigma}}(\nu)\, \vec{E}(\vec{r}, \nu) \tag{39c}$$

with $\vec{B}$, $\vec{H}$, $\vec{D}$, $\vec{E}$, and $\vec{J}$ all two-dimensional vectors in the polarization vector space,

$$\underline{\underline{\mu}}(\nu) = \begin{pmatrix} \mu_{11}(\nu) & \mu_{12}(\nu) \\ \mu_{21}(\nu) & \mu_{22}(\nu) \end{pmatrix}; \quad \underline{\underline{\epsilon}}(\nu) = \begin{pmatrix} \epsilon_{11}(\nu) & \epsilon_{12}(\nu) \\ \epsilon_{21}(\nu) & \epsilon_{22}(\nu) \end{pmatrix}; \quad \underline{\underline{\sigma}}(\nu) = \begin{pmatrix} \sigma_{11}(\nu) & \sigma_{12}(\nu) \\ \sigma_{21}(\nu) & \sigma_{22}(\nu) \end{pmatrix}$$

Maxwell's equations become

$$\nabla \times \vec{E}(\vec{r}, \nu) = -i\omega\underline{\underline{\mu}}(\nu)\, \vec{H}(\vec{r}, \nu) \tag{40a}$$

$$\nabla \times \vec{H}(\vec{r}, \nu) = [\underline{\underline{\sigma}}(\nu) + i\omega\underline{\underline{\epsilon}}(\nu)]\, \vec{E}(\vec{r}, \nu). \tag{40b}$$

Here, $\underline{\underline{\sigma}}(\nu)$ and $\underline{\underline{\epsilon}}(\nu)$ may always be taken as Hermitian. Any complex matrix, C, has a unique decomposition

$$\underline{\underline{C}} = \underline{\underline{C}}_1 + i\underline{\underline{C}}_2, \tag{41}$$

where both $\underline{\underline{C}}_1$, and $\underline{\underline{C}}_2$ are Hermitian. This is the matrix equivalent to separating a complex number into real and imaginary parts.

If $\underline{\underline{\mu}}$ and $\underline{\underline{\epsilon}}$ are independent of polarization, then

$$\underline{\underline{\mu}}(\nu) = \mu_0\underline{\underline{I}} \tag{42a}$$

$$\underline{\underline{\epsilon}}(\nu) = \epsilon_0\underline{\underline{I}}. \tag{42b}$$

From Eqs. 40, 42, and the TEM assumption the wave equations follow.

$$\frac{d^2}{dz^2}\, \vec{E}(\vec{r}, \nu) = i\omega\mu_0(\underline{\underline{\sigma}}(\nu) + i\omega\epsilon_0\underline{\underline{I}})\, \vec{E}(\vec{r}, \nu) \tag{43a}$$

$$\frac{d^2}{dz^2}\, \vec{H}(\vec{r}, \nu) = i\omega\mu_0(\underline{\underline{\sigma}}(\nu) + i\omega\epsilon_0\underline{\underline{I}})\, \vec{H}(\vec{r}, \nu), \tag{43b}$$

where z is the direction of propagation of the wave.

The solution to (43a) is sought. This is of the form

$$\frac{d^2}{dz^2}\vec{E}(z,\nu) = \underline{\underline{B}}(\nu)\,\vec{E}(z,\nu) \tag{44a}$$

$$\underline{\underline{B}}(\nu) = -\omega^2\mu_o\epsilon_o\underline{I} + i\omega\mu_o\underline{\underline{\sigma}}(\nu). \tag{44b}$$

Let

$$\underline{\underline{A}}^2(\nu) = \underline{\underline{B}}(\nu). \tag{44c}$$

Then

$$\vec{E}(z,\nu) = e^{-\underline{\underline{A}}(\nu)\Delta z}\,\vec{E}(z_o,\nu) \tag{45}$$

is a solution to (44a). Hence the task is to find $\underline{\underline{A}}(\nu)$. (The same expression with plus sign in the exponential is also a solution, but represents a negative z-traveling wave.)

Decompose $\underline{\underline{A}}(\nu)$ into its form as in (5):

$$\underline{\underline{A}}(\nu) = \underline{\underline{A}}_1(\nu) + i\underline{\underline{A}}_2(\nu) \tag{46}$$

with $\underline{\underline{A}}_1(\nu)$ and $\underline{\underline{A}}_2(\nu)$ Hermitian. Then from (44b) and (44c),

$$\underline{\underline{A}}_2^2 - \underline{\underline{A}}_1^2 = \omega^2\mu_o\epsilon_o\underline{I} \tag{47a}$$

$$\underline{\underline{A}}_1\underline{\underline{A}}_2 + \underline{\underline{A}}_2\underline{\underline{A}}_1 = \omega\mu_o\underline{\underline{\sigma}}. \tag{47b}$$

$A_1(\nu)$ is now seen to be the matrix equivalent to the attenuation constant, and $\underline{\underline{A}}_2(\nu)$ the matrix equivalent of the propagation constant. The small-loss assumption (tr $\underline{\underline{\sigma}}(\nu) \ll \omega\epsilon_o$) yields

$$\underline{\underline{A}}_2(\nu) \approx \omega\sqrt{\mu_o\epsilon_o}\,\underline{I} \tag{48a}$$

$$\underline{\underline{A}}_1(\nu) \approx \frac{1}{2}\sqrt{\frac{\mu_o}{\epsilon_o}}\,\underline{\underline{\sigma}}(\nu). \tag{48b}$$

Equation 48a states that all polarizations have the same propagation velocity; whereas Eq. 48b states that the attenuation (absorption) can be polarization-dependent. Further investigation will also show that

$$|\vec{H}(z,\nu)| = \sqrt{\frac{\epsilon_o}{\mu_o}}\,|\vec{E}(z,\nu)|, \tag{49}$$

as might be expected.

Note that if N incoherent absorption processes are occurring simultaneously, then

the over-all $\underline{\underline{A}}_1(\nu)$ is given by

$$\underline{\underline{A}}_{1_{total}}(\nu) = \sum_{n=1}^{N} \underline{\underline{A}}_{1_n}(\nu),$$

(50)

as is seen by looking once again at Maxwell's equations.

Since the $\underline{\underline{A}}_{1_{total}}(\nu)$ is Hermitian, there is a polarization basis in which it is diagonal (no "crosstalk" between polarizations), and experiments can be most readily performed.

W. B. Lenoir

References

1. W. B. Davenport and W. L. Root, Random Signals and Noise (McGraw-Hill Book Company, New York, 1958).

2. M. Born and E. Wolf, Principles of Optics (Pergamon Press, Oxford, 1959).

3. H. C. Ko, On the reception of quasi-monochromatic, partially polarized radio waves, Proc. IRE 50, 1950-1957 (1962).

4. S. Chandrasekhar, Radiative Transfer (Dover Publications, New York, 1960).

D. FIVE-MILLIMETER RADIATIVE TRANSFER IN THE UPPER ATMOSPHERE

1. Introduction

The O_2 molecule in the ground state has no electric dipole moment, but does have a magnetic dipole moment resulting from the unpaired spins of two electrons.[1] This magnetic moment permits microwave transitions between the 5 structure levels of the molecular rotational states.

The electron-spin quantum number, S, is equal to 1 for the O_2 molecule. The total angular momentum quantum number, J, is given by

$$J = N + 1$$

$$J = N$$

(1)

$$J = N - 1,$$

where N, the rotational quantum number, must be odd because of the exclusion principle. The selection rules permit the transitions

$$J = N \rightarrow J = N + 1,$$

called the N^+ transition, and

$$J = N \rightarrow J = N - 1,$$

called the N^- transition.

If there is no external magnetic field, the radiation will be isotropic and unpolarized. The introduction of an external magnetic field will greatly complicate the picture by introducing radiation that is neither isotropic nor unpolarized.

2. Zeeman Splitting and Matrix Values

The application of an external magnetic field will cause a splitting of the resonance lines because the magnetic moment associated with the O_2 molecule couples with the external field to split the energy associated with a given J into $2J + 1$ levels corresponding to $M = -J, \ldots 0, \ldots J$, where M is the quantum number associated with the projection of the O_2 magnetic moment along the direction of the external field. This perturbation[1] is given by

$$\Delta W = -1.001\mu_o MH \frac{J(J+1) + S(S+1) - N(N+1)}{J(J+1)} \tag{2}$$

in which μ_o is the Bohr magneton, H is the external field strength, and $S = 1$.

The selection rules permit transitions in J to be accompanied by changes in M of $\Delta M = 0, \pm 1$. The transition frequencies for the case of no external magnetic field are well known.[2] The change in this frequency, $\Delta\nu$, caused by an external magnetic field is given in Table III-4.

Table III-4. Frequency change.

$$\Delta M = M_{final} - M_{initial}$$

$k = 2.8026$ for $\Delta\nu$ in Mc/sec

	$J = N \rightarrow J = N + 1$	$J = N \rightarrow J = N - 1$
$\Delta M = +1$	$\dfrac{kH}{N+1}\left(1 + M\dfrac{N-1}{N}\right)$	$-\dfrac{kH}{N}\left(1 + M\dfrac{N+2}{N+1}\right)$
$\Delta M = 0$	$\dfrac{kH}{N+1} M \dfrac{N-1}{N}$	$-\dfrac{kH}{N} M \dfrac{N+2}{N+1}$
$\Delta M = -1$	$\dfrac{kH}{N+1}\left(-1 + M\dfrac{N-1}{N}\right)$	$-\dfrac{kH}{N}\left(-1 + M\dfrac{N+2}{N+1}\right)$

The corresponding matrix elements are readily found by using the tabulated relative intensities[3] and the fact that for $H = 0$ the total radiation associated with a given transition must be isotropic, unpolarized, and equal in strength to the analysis working directly with the non-Zeeman matrix elements.[1] Table III-5 shows the matrix elements.

Table III-5. Matrix elements, $|\mu|^2$.

	$J = N \rightarrow J = N + 1$	$J = N \rightarrow J = N - 1$
$\Delta M = +1$	$\dfrac{3}{2} \dfrac{N(N+M+1)(N+M+2)}{(N+1)^2 (2N+1)} \mu_o^2$	$\dfrac{3}{2} \dfrac{(N+1)(N-M)(N-M-1)}{N^2(2N+1)} \mu_o^2$
$\Delta M = 0$	$3 \dfrac{N[(N+1)^2 - M^2]}{(N+1)^2 (2N+1)} \mu_o^2$	$3 \dfrac{(N+1)(N^2 - M^2)}{N^2(2N+1)} \mu_o^2$
$\Delta M = -1$	$\dfrac{3}{2} \dfrac{N(N-M+1)(N-M+2)}{(N+1)^2 (2N+1)} \mu_o^2$	$\dfrac{3}{2} \dfrac{(N+1)(N+M)(N+M-1)}{N^2(2N+1)} \mu_o^2$

3. Complex Propagation Matrix

In this situation the velocity of propagation is independent of polarization. Furthermore, for brightness-temperature coherency spectrum matrix calculations the only part of $\underline{\underline{A}}(\nu)$ that matters is its Hermitian part, since the anti-Hermitian part (i times a Hermitian matrix) is a constant times the unit matrix, and this will not appear in the matrix solution (see Secs. III-B and III-C).

For a given J transition ($J=N \rightarrow J=N \pm 1$) there are three unrelated processes, those for $\Delta M = 0$, $\Delta M = +1$, $\Delta M = -1$. Hence the $\underline{\underline{A}}_{total}(\nu)$ will be the sum of the individual $\underline{\underline{A}}(\nu)$. The radiation emitted from any one of these processes is totally polarized.[3] The type of polarization depends on the direction of observation (that is, on the angle between the external field and the observing direction).

Since the radiation from one process is totally polarized, the medium will be transparent to the opposite polarization. On this polarization basis

$$\underline{\underline{A}}_{\Delta M}(\nu) = \begin{pmatrix} a_{\Delta M}(\nu) & 0 \\ 0 & 0 \end{pmatrix} \tag{3}$$

in which $\underline{\underline{A}}_2(\nu)$ has been neglected. This polarization basis will be different for each of the processes ($\Delta M = 0, \pm 1$). On this basis, it is easy to discuss an experiment to measure $\underline{\underline{A}}_{\Delta M}(\nu)$, inasmuch as it is now a scalar problem.

The $\underline{\underline{A}}_{\Delta M}(\nu)$ obtained can be transformed to a common basis and summed.

$$\underline{\underline{A}}(\nu) = \underline{\underline{A}}_{\Delta M=+1}(\nu) + \underline{\underline{A}}_{\Delta M=0}(\nu) + \underline{\underline{A}}_{\Delta M=-1}(\nu), \tag{4}$$

where the $\underline{\underline{A}}_{\Delta M}(\nu)$ must all be on the same basis.

The chosen standard basis is convenient for the problem of the reception of microwave radiation from the atmosphere by an earth-orbiting satellite. Let R, θ, ϕ be the normal spherical coordinates (center of earth as origin, north rotation axis as +z) of

the observation point. The direction of observation is in the $-\vec{L}_R$ direction. Let the polarization basis be

α, linear polarization in $\vec{L}_\theta$ direction.

β, linear polarization in $\vec{L}_\phi$ direction.

The magnetic North Pole and the rotational North Pole do not coincide. Hence, there will in general be a ϕ component of the magnetic field which would not be as it is if the two poles coincided.

At a given point along the observational path, r, θ, ϕ with $r \leqslant R$, let the angle between B, the magnetic field, and R define ψ_{BP}; $\psi_{BP} \mathscr{E} [0, \pi]$. Also, let the angle between the projections of B and $\vec{L}_z$ on a plane perpendicular to R define ψ_{BN}; $\psi_{BN} \mathscr{E} [-\pi, \pi]$. ψ_{BN} is a consequence of the two poles not coinciding.

In this coordinate system each of the $\underline{\underline{A}}_{\Delta M}(\nu)$ will be of the form

$$
\begin{pmatrix}
\rho_{11}(\nu) & \rho_{12_R}(\nu) + i\rho_{12_I}(\nu) \\
\rho_{12_R}(\nu) - i\rho_{12_I}(\nu) & \rho_{22}(\nu)
\end{pmatrix}
\tag{5}
$$

with the 1 coordinate being the θ-directed linear polarization (for $\vec{\mathscr{e}}$), and 2 being the ϕ-directed linear polarization.

The $\underline{\underline{A}}_{\Delta M}(\nu)$ for one process[2,4] is given by

$$
\underline{\underline{A}}_{\Delta M}(\nu) = c \frac{p\nu^2}{T^3} e^{-E_J/T} \underline{\underline{\rho}}_{\Delta M} \sum_{M=-J}^{J} |\mu_{J, \Delta M}|^2 F(\nu, \nu_{J, \Delta M}, \Delta\nu_d, \Delta\nu_c),
\tag{6}
$$

in which

p = pressure in mm Hg

T = temperature in degrees Kelvin

ν = frequency in Gc/sec

E_J = energy of Jth level in degrees Kelvin

$|\mu|^2$ = matrix elements listed above

c = 0.30506, a constant for $\underline{\underline{A}}$ having units of km^{-1}.

$F(\nu, \nu_{J, \Delta M}, \Delta\nu_d, \Delta\nu_c)$ is the convolution of the pressure-broadening (Lorentz) shape with the Doppler (gauss) shape for $\Delta\nu_d$ = Doppler half-width, $\Delta\nu_c$ = collision half-width, and $\nu_{J, \Delta M}$ = split resonant-frequency component.

$\underline{\underline{\rho}}_{\Delta M}$ is a Hermitian matrix describing the angular (ψ_{BP} and ψ_{BN}) dependence of polarization. It is of the form

$$
\underline{\underline{\rho}}_{\Delta M} =
\begin{pmatrix}
\rho_{11} & \rho_{12_R} + i\rho_{12_I} \\
\rho_{12_R} - i\rho_{12_I} & \rho_{22}
\end{pmatrix}.
\tag{7}
$$

For $\Delta M = 0$,

$$\rho_{11} = \sin^2 \psi_{BP} \sin^2 \psi_{BN}$$

$$\rho_{12_R} = \sin^2 \psi_{BP} \sin \psi_{BN} \cos \psi_{BN}$$

$$\rho_{12_I} = 0$$

$$\rho_{22} = \sin^2 \psi_{BP} \cos^2 \psi_{BN}.$$

(8)

For $\Delta M = \pm 1$,

$$\rho_{11} = \frac{1}{2}\left[\cos^2 \psi_{BN} + \sin^2 \psi_{BN} \cos^2 \psi_{BP} \right]$$

$$\rho_{12_R} = -\frac{1}{2} \sin^2 \psi_{BP} \sin \psi_{BN} \cos \psi_{BN}$$

$$\rho_{12_I} = \mp\frac{1}{2} \cos \psi_{BP}$$

$$\rho_{22} = \frac{1}{2}\left[\sin^2 \psi_{BN} + \cos^2 \psi_{BN} \cos^2 \psi_{BP} \right].$$

(9)

The complex propagation constant is completed by summing over the possible J transitions.

4. Radiative Transfer Solution

The problem is to solve

$$\frac{d}{dz} \underline{\underline{T}}_B(\nu) + \underline{\underline{A}}(\nu) \underline{\underline{T}}_B(\nu) + \underline{\underline{T}}_B(\nu) \underline{\underline{A}}^{t*}(\nu) = 2t_k \underline{\underline{A}}(\nu),$$

(10)

where t_k is the atmospheric kinetic temperature (a function of z), and the Hermitian nature of $\underline{\underline{A}}(\nu)$ has been used. The method is to approximate the atmosphere as a series of constant-temperature, constant-pressure layers, each 1 km thick, extending from the ground to a height of 100 km.

The solution to Eq. 10 for the transfer through one such layer is

$$\underline{\underline{T}}_{B_o}(\nu) = e^{-\underline{\underline{A}}(\nu)\Delta z} \underline{\underline{T}}_{B_i}(\nu) e^{-\underline{\underline{A}}(\nu)\Delta z} + t_k[\underline{\underline{I}} - e^{-2\underline{\underline{A}}(\nu)\Delta z}],$$

(11)

where $\underline{\underline{T}}_{B_i}(\nu)$ is the brightness-temperature coherency spectrum matrix incident on the layer, t_k is the temperature of the layer, and $\underline{\underline{A}}(\nu)$ is the complex propagation
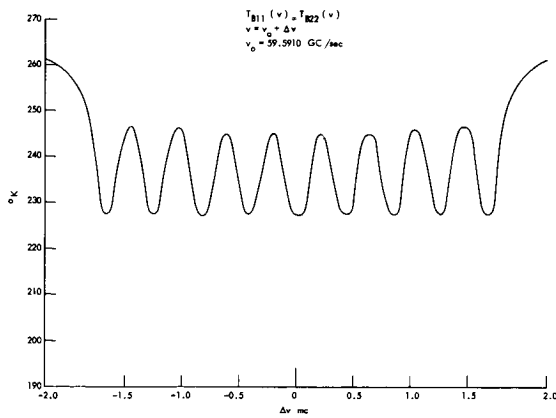
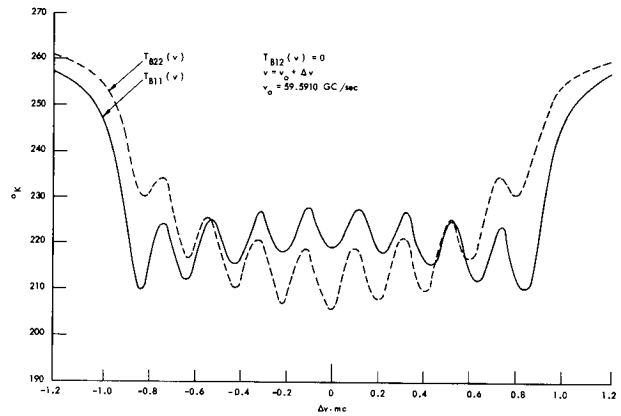Fig. III-7. J = 5 → J = 4 transition
(magnetic pole).

Fig. III-9. J = 5 → J = 4 transition
(magnetic equator).

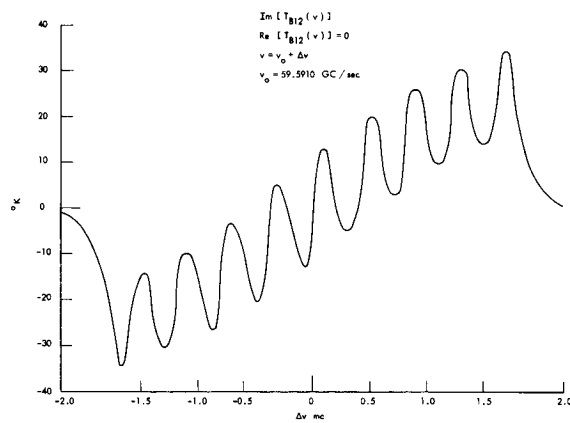Fig. III-8. J = 5 → J = 4 transition
(magnetic pole).

matrix of the layer.

Continued application of Eq. 11 yields the $\underline{\underline{T}}_B(\nu)$ as received at the satellite. A slightly different analysis emphasizing the contribution to $\underline{\underline{T}}_B(\nu)$ from the various heights would involve the matrix weighting functions. In this procedure the emission from one layer is treated from emission height to satellite. Each height is treated accordingly, and the results summed.

The weighting function matrix is found to be

$$\underline{\underline{WF}}(h_i, \nu) = \underline{\underline{P}}(h_i, \nu) \left[I - e^{-2\underline{\underline{A}}(\nu_s, h_i)\,\Delta h_i} \right] \underline{\underline{P}}^{t*}(h_i, \nu) \tag{12}$$

with

$$\underline{\underline{P}}(h_i, \nu) = e^{-\underline{\underline{A}}(\nu, h_s)\,\Delta h_s} \; e^{-\underline{\underline{A}}(\nu, h_{s-1})\,\Delta h_{s-1}} \; \ldots \; e^{-\underline{\underline{A}}(\nu, h_{i+1})\,\Delta h_{i+1}}. \tag{13}$$

The $\underline{\underline{T}}_B(\nu)$ is obtained from (12) by

$$\underline{\underline{T}}_B(\nu) = \sum_{h_i = \text{ground}}^{\text{satellite}} WF(h_i, \nu)\, t(h_i). \tag{14}$$

Examples of the $\underline{\underline{T}}_B(\nu)$ computed for two resonance lines at positioins corresponding to both the magnetic pole and the magnetic equator are presented in Figs. III-7, III-8, and III-9.

A particular point on the magnetic equator was chosen so that $T_{B_{12}}(\nu) = 0$. The fine structure resulting from the individual Zeeman components is easily seen. A dipole model of the earth's magnetic field was used, with $|B| = 0.624$ gauss at the pole.

W. B. Lenoir

References

1. C. H. Townes and A. L. Schawlow, Microwave Spectroscopy (McGraw-Hill Book Company, New York, 1955).

2. M. L. Meeks and A. E. Lilley, The microwave spectrum of oxygen in the Earth's atmosphere, J. Geophys. Res. 68, 1683-1703 (1963).

3. E. U. Condon and G. H. Shortley, The Theory of Atomic Spectra (Cambridge University Press, London, 1951).

4. J. H. VanVleck and V. F. Weisskopf, On the shape of collision-broadened lines, Rev. Modern Phys. 17, 3 (1945).

IV. OPTICAL AND INFRARED SPECTROSCOPY

Prof. C. H. Perry   W. J. Hakel   P. Lubitz
Dr. R. Geick   D. B. Hall   E. C. Reifenstein III
D. P. Athans   E. F. Young

A. ANALYSIS OF REFLECTION SPECTRA BY MEANS OF A FIT WITH A CLASSICAL DISPERSION FORMULA

In the infrared region polar crystals exhibit a reflection spectrum with reststrahlen bands because of the infrared active lattice modes. These modes, which have a linear dipole moment with respect to the normal coordinate, cause the infrared dispersion. When the reflectivity of such a crystal has been measured over a sufficiently wide frequency range, two kinds of analyses are generally used to extract the available information, such as eigenfrequencies, oscillator strengths, and so forth. One method is the Kramers-Kronig analysis,[1] and the other is an optimum fit of the data by means of a classical dispersion formula. The latter method will be discussed in this report with respect to some theoretical aspects and its practical application.

Polar and cubic crystals with two particles per unit cell have one infrared active lattice mode. The classical treatment of the interaction of electromagnetic waves and this lattice mode yields the classical dispersion formula[2] for the complex dielectric constant of the crystal as a function of the frequency ω

$$\epsilon = \epsilon' - i\epsilon'' = \epsilon_\infty + \frac{s^2}{\omega_o^2 - \omega^2 + i\omega\gamma}, \tag{1}$$

where ϵ_∞ is the contribution to ϵ of the UV-absorption and UV-dispersion which can be assumed to be constant in the infrared region; ω_o is the eigenfrequency of the infrared active mode; s^2 is its oscillator strength; and γ is the damping constant.

If a more complicated crystal is considered with more than one infrared active mode that do not interact with each other, the dispersion formula contains a sum of dispersion terms, one for each active mode. In certain cases it may be necessary to take the interaction of at least two modes into account, and Barker and Hopfield[3] have derived a dispersion formula for the case in which two modes strongly interact. In the equations of motion an interaction term was added:

$$\ddot{\vec{y}}_1 + \rho_{11}\dot{\vec{y}}_1 + a_{11}\vec{y}_1 + a_{12}\vec{y}_2 = a_{13}\vec{E}$$

$$\ddot{\vec{y}}_2 + \rho_{22}\dot{\vec{y}}_2 + a_{21}\vec{y}_1 + a_{22}\vec{y}_2 = a_{23}\vec{E}. \tag{2}$$

The transformation diagonalizing the force constant matrix

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

yields off-diagonal terms for the damping constants:

$$\ddot{\vec{x}}_1 + \gamma_{11}\dot{\vec{x}}_1 + \gamma_{12}\dot{\vec{x}}_2 + \omega_1^2\vec{x}_1 = b_{13}\vec{E}$$

$$\ddot{\vec{x}}_2 + \gamma_{21}\dot{\vec{x}}_1 + \gamma_{22}\dot{\vec{x}}_2 + \omega_2^2\vec{x}_2 = b_{23}\vec{E}. \tag{3}$$

From these two equations the following dispersion formula in which a third dispersion term is added for a third infrared active eigenvibration that does not interact with the two others[4] is obtained:

$$\epsilon = \epsilon' - i\epsilon'' = \epsilon_\infty + \frac{s_1^2\left(\omega_2^2-\omega^2+i\omega\gamma_2\right) - 2is_1s_2\omega\gamma_{12} + s_2^2\left(\omega_1^2-\omega^2+i\omega\gamma_1\right)}{\left(\omega_1^2-\omega^2+i\omega\gamma_1\right)\left(\omega_2^2-\omega^2+i\omega\gamma_2\right) + \omega^2\gamma_{12}^2} + \frac{s_3^2}{\omega_3^2 - \omega^2 + i\omega\gamma_3}, \tag{4}$$

where the subscripts 1, 2, and 3 refer to the three active modes, and γ_{12} is the damping term arising from the interaction of modes 1 and 2.

Since the damping constants γ_ν and $\gamma_{\nu\mu}$ are introduced ad hoc in the classical treatment, this derivation of a dispersion formula cannot show the physical meaning of the damping terms. The results of a proper quantum mechanical treatment of the infrared dispersion, however, will show the damping related to the dissipation of energy from the active lattice modes (dispersion oscillators) to other lattice modes by means of anharmonic terms in the lattice potential.[5] If third-order anharmonic potential only is taken into account and all nonlinear terms in the dipole moment are neglected, the following expression holds for the dielectric constant of a cubic crystal with two infrared active lattice modes[6]

$$\epsilon = \epsilon_\infty + \frac{s_1^2\left(\omega_2^2-\omega^2+i\delta_2\right) - 2is_1s_2\delta_{12} + s_2^2\left(\omega_1^2-\omega^2+i\delta_1\right)}{\left(\omega_1^2-\omega^2+i\delta_1\right)\left(\omega_2^2-\omega^2+i\delta_2\right) + \delta_{12}^2}. \tag{5}$$

This is the same formula that is obtained by the classical treatment, but instead of damping constants γ there are frequency-dependent damping functions δ:

$$\delta_j = \text{const} \sum_{j'j''} \int \left| \Phi\begin{pmatrix} 0 & k & -k \\ j & j' & j'' \end{pmatrix}\right|^2 \frac{\left[n'+\frac{1}{2}\right] \pm \left[n''+\frac{1}{2}\right]}{\omega'\omega''} \Delta(\omega\pm\omega'-\omega'') \, dk^3$$

$$\delta_{j\ell} = \delta_{\ell j} = \text{const} \sum_{j'j''} \int \Phi\begin{pmatrix} 0 & k & -k \\ j & j' & j'' \end{pmatrix} \Phi\begin{pmatrix} 0 & k & -k \\ \ell & j' & j'' \end{pmatrix} \frac{\left[n' + \frac{1}{2}\right] \pm \left[n'' + \frac{1}{2}\right]}{\omega'\omega''} \Delta(\omega \pm \omega' - \omega'') \, dk^3,$$

(6)

where $\Phi\begin{pmatrix} 0 & k & -k \\ j & j' & j'' \end{pmatrix}$ is the third-order potential coefficient coupling modes $\begin{pmatrix} 0 \\ j \end{pmatrix}$ [wave

vector 0, branch j], $\begin{pmatrix} k \\ j' \end{pmatrix}$ and $\begin{pmatrix} -k \\ j'' \end{pmatrix}$; ω', ω'', n', and n'' are frequencies and the thermal

averages of phonon occupation numbers of modes $\begin{pmatrix} k \\ j' \end{pmatrix}$ and $\begin{pmatrix} -k \\ j'' \end{pmatrix}$, respectively; Δ refers

to the Dirac delta function; and the integrals are to be taken over the volume of the first

Brillouin zone in reciprocal space.

Essentially, the damping functions δ_1 and δ_2 are the probabilities for all the 2-phonon summation processes (+ sign) or 2-phonon difference processes (- sign) via the dispersion oscillators 1 and 2 which are consistent with wave vector and energy conservation at a given photon frequency ω. δ_{12} is the interference of the processes contained in δ_1 and δ_2.

For our purpose, the analysis of a reflection spectrum is carried out by means of the classical formula with constant damping terms. The reflection spectrum exhibits only the main features of the infrared dispersion and not the details of the absorption spectrum, especially outside the reststrahlen bands where the reflectivity is given mainly by the refraction index; even in the reststrahlen bands it is not possible to evaluate more than one damping function from the experimental data. Consequently, the use of the classical formula may be justified and the damping constants should be understood to be an average of the damping functions, mainly in the neighborhood of the eigenfrequencies. In the actual analysis, the constants in Eq. 4 have to be chosen by trial and error, and then ϵ is evaluated for the frequency range under consideration. From ϵ' and ϵ'' the optical constants n and k are obtained by means of the equations

a) $\quad \epsilon' > 0 \qquad n = \sqrt{\frac{1}{2}\left(\epsilon' + \sqrt{\epsilon'^2 + \epsilon''^2}\right)} \qquad k = \frac{\epsilon''}{2n}$

b) $\quad \epsilon' = 0 \qquad n = k = \sqrt{\frac{\epsilon''}{2}}$

(7)

c) $\quad \epsilon' < 0 \qquad k = \sqrt{\frac{1}{2}\left(-\epsilon' + \sqrt{\epsilon'^2 + \epsilon''^2}\right)} \qquad n = \frac{\epsilon''}{2k}$

The reflectivity (given only for normal incidence)

$$R = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} \tag{8}$$

can be evaluated and compared with experimental data. By variation of the constants in Eq. 4, an optimum fit to the experimental data is obtained, and this analysis yields the eigenfrequencies, the oscillator strengths, and the damping constants of the infrared active modes.

Note that Eq. 4 can be used also in the case of a crystal having two interacting eigenvibrations and additional dispersion and absorption caused by free carriers (for example, semiconductors). By setting $\omega_3 = 0$, the last dispersion term in Eq. 4 is converted into a Drude-term, and s_3 and γ_3 are determined by the carrier concentration N, the effective mass m^*, and the mobility μ of the free carriers

$$s_3 = \sqrt{\frac{N}{m^* \epsilon_{vac}}} \frac{e}{2\pi c} \; [cm^{-1}] \qquad \gamma_3 = \frac{e}{2\pi c m^* \mu} [cm^{-1}],$$

where e is the electron charge, c is the light velocity in vacuum, and ϵ_{vac} is the dielectric constant in vacuum.

As examples of this analysis we shall describe the fit to the reflection spectra of BN and $KNiF_3$. BN is a uniaxial anisotropic crystal with a layer structure similar to graphite. The reflectivity has been measured with linearly polarized light with the E-vector parallel and perpendicular to the c-axis (See Figs. IV-1 and IV-2). $KNiF_3$ has the perovskite structure and its experimentally determined reflectivity is shown in Fig. IV-3. From the experimental data, it appears that there are two infrared active lattice modes causing reststrahlen bands in both cases for BN, and three active modes in $KNiF_3$. The initial values for the various constants in the classical dispersion formula (Eq. 4) were obtained from the results of a Kramers-Kronig analysis (see Figs. IV-4, IV-5 and IV-6).

BN   E ∥ c

$\epsilon_\infty = 3.06$	$s_1 = 771$	$\omega_1 = 783$	$\gamma_1 = 15.0$	(10a)
$\gamma_{12} = 0$	$s_2 = 1585$	$\omega_2 = 1510$	$\gamma_2 = 146$	
$\epsilon_0 = 5.13$	$s_3 = 0$	$\omega_3 = 0$	$\gamma_3 = 0$	

BN   E ⊥ c

$\epsilon_\infty = 5.36$	$s_1 = 351$	$\omega_1 = 767$	$\gamma_1 = 35.0$	(10b)
$\gamma_{12} = 0$	$s_2 = 1850$	$\omega_2 = 1367$	$\gamma_2 = 31.0$	
$\epsilon_0 = 7.45$	$s_3 = 0$	$\omega_3 = 0$	$\gamma_3 = 0$	

<u>KNiF$_3$</u>   (isotropic crystal)

$\epsilon_\infty = 2.47$	$s_1 = 233$	$\omega_1 = 152$	$\gamma_1 = 19.5$	
$\gamma_{12} = 0$	$s_2 = 355$	$\omega_2 = 245$	$\gamma_2 = 15.0$	(10c)
$\epsilon_0 = 8.13$	$s_3 = 483$	$\omega_3 = 447$	$\gamma_3 = 32.5$	

The units for s_ν, ω_ν, γ_ν and $\gamma_{\mu\nu}$ are in cm^{-1}, and ϵ_∞ is the dielectric constant relative to that of vacuum. The reflectivity computed with these sets of constants has to be compared with the experimental data (see Figs. IV-1, IV-2 and IV-3), and the agreement at first is still rather poor for most of the reststrahlen bands. The computed values of ϵ' and ϵ'' can be compared with the Kramers-Kronig data (see Figs. IV-4, IV-5 and IV-6). In order to improve the fit, it is often advisable to make use of the following rules.

1. The static dielectric constant which determines the reflectivity at the low-frequency end of the spectrum outside the reststrahlen bands is given by

$$\epsilon_0 = \epsilon_\infty + \frac{s_1^2}{\omega_1^2} + \frac{s_2^2}{\omega_2^2} + \frac{s_3^2}{\omega_3^2}$$

2. Where the real part ϵ' changes from positive to negative values, the reflectivity exhibits a pronounced minimum and a sharp rise to high values with decreasing frequencies, provided ϵ'' is sufficiently small.

3. The height of a reststrahlen band is determined by ϵ'', that is, mainly by the damping γ_ν, while the interaction damping $\gamma_{\mu\nu}$ becomes significant in the region between two eigenvibrations.

In this way, the best fit for BN was obtained with the following sets of constants:

<u>BN   E $\parallel$ c</u>

$\epsilon_\infty = 4.10$	$s_1 = 572$	$\omega_1 = 783$	$\gamma_1 = 8.0$	
$\gamma_{12} = 0$	$s_2 = 1020$	$\omega_2 = 1510$	$\gamma_2 = 80.0$	(11a)
$\epsilon_0 = 5.09$	$s_3 = 0$	$\omega_3 = 0$	$\gamma_3 = 0$	

<u>BN   E $\perp$ c</u>

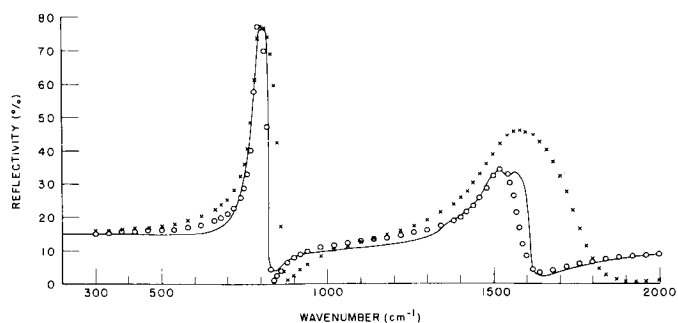|				
$\epsilon_\infty = 4.95$	$s_1 = 351$	$\omega_1 = 767$	$\gamma_1 = 35.0$	
$\gamma_{12} = 0$	$s_2 = 1870$	$\omega_2 = 1367$	$\gamma_2 = 29.0$	(11b)
$\epsilon_0 = 7.04$	$s_3 = 0$	$\omega_3 = 0$	$\gamma_3 = 0$	

Fig. IV-1. Reflectivity of BN, E ∥ c. Experimental (———), computed by using the classical dispersion formula and the values for the constants given in (10a) (X X X) and (11a) (o o o).
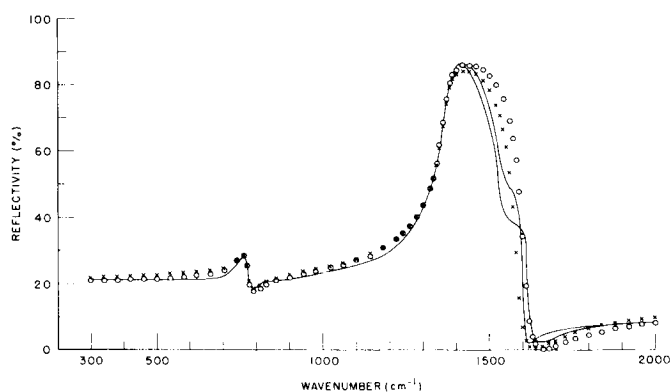

Fig. IV-2. Reflectivity of BN, E ⊥ c. Experimental (———), computed by using the classical dispersion formula and the values for the constants given in (10b) (X X X) and (11b) (o o o).
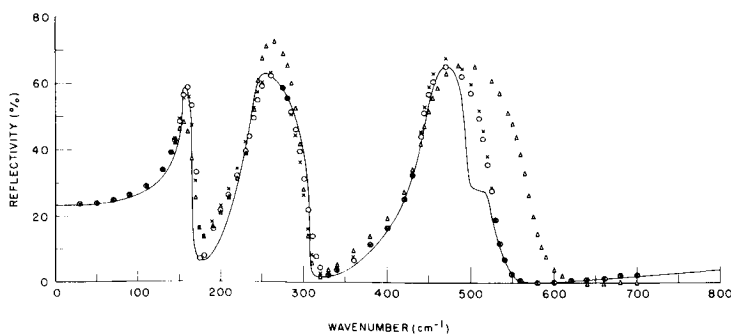

Fig. IV-3. Reflectivity of KNiF$_3$. Experimental (———), computed by using the classical dispersion formula and the values for the constants given in (10c) (Δ Δ Δ), (12) (X X X) and (13) (o o o).

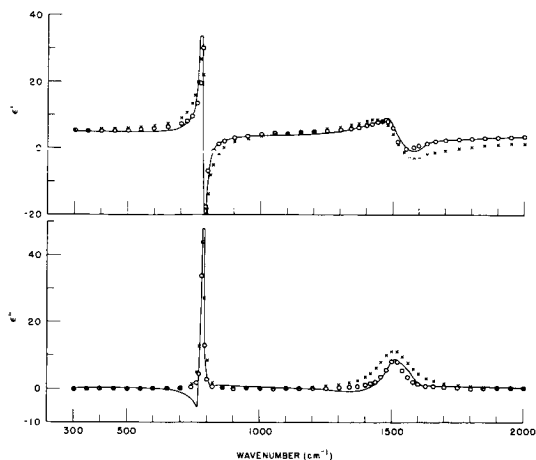Fig. IV-4. Real part ϵ' and imaginary part ϵ'' of the dielectric constant of BN for E ‖ c obtained by Kramers-Kronig analysis (———), computed by using the classical dispersion formula and the values for the constants given in (10a) (X X X) and (11a) (o o o).
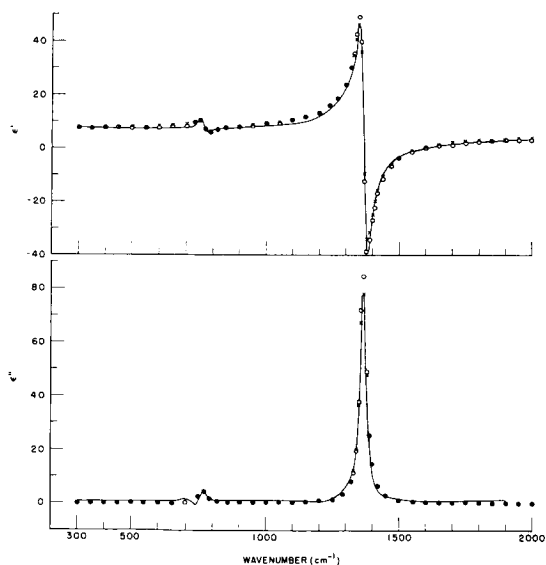

Fig. IV-5. Real part ϵ' and imaginary part ϵ'' of the dielectric constant of BN for E ⊥ c obtained by Kramers-Kronig analysis (———), computed by using the classical dispersion formula and the values for the constants given in (10b) (X X X) and (11b) (o o o).
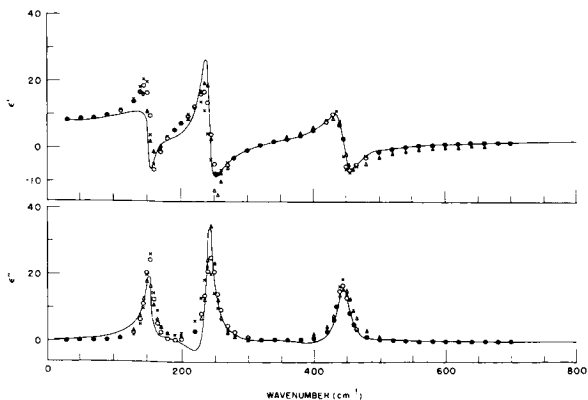

Fig. IV-6. Real part ϵ' and imaginary part ϵ'' of the dielectric constant of $KNiF_3$ obtained by Kramers-Kronig analysis (————), computed by using the classical dispersion formula and the values for the constants given in (10c) (Δ Δ Δ), (12) (X X X) and (13) (o o o).

In these cases it was not necessary to use the interaction damping. The computed data are shown in detail in Figs. IV-1, IV-2, IV-4 and IV-5.

For $KNiF_3$ the best fit without interaction damping was obtained with the following

$$\epsilon_\infty = 2.85 \quad s_1 = 230 \quad \omega_1 = 154 \quad \gamma_1 = 13.0$$
$$\gamma_{12} = 0 \quad s_2 = 365 \quad \omega_2 = 241 \quad \gamma_2 = 25.0 \quad (12)$$
$$\epsilon_0 = 8.21 \quad s_3 = 408 \quad \omega_3 = 444 \quad \gamma_3 = 20.0$$

The agreement of computed and experimental data is quite reasonable in the region of the reststrahlen bands (Fig. IV-3) but the computed reflectivity is too high in the frequency region 170-220 cm^{-1}. Therefore better agreement is reached with $\gamma_{12} \neq 0$ (Fig. IV-3).

$$\epsilon_\infty = 2.85 \quad s_1 = 230 \quad \omega_1 = 154 \quad \gamma_1 = 13.5$$
$$\gamma_{12} = 12.0 \quad s_2 = 371 \quad \omega_2 = 245 \quad \gamma_2 = 22.5 \quad (13)$$
$$\epsilon_0 = 8.21 \quad s_3 = 408 \quad \omega_3 = 444 \quad \gamma_3 = 22.0$$

The computed values of ϵ' and ϵ'' are shown in Fig. IV-6.

The computations for this work were performed at the Computation Center, M.I.T., and the authors wish to thank S. A. Rappaport of the RLE Computation Group for writing the program.

R. Geick, C. H. Perry

References

1. C. H. Perry, Quarterly Progress Report No. 71, Research Laboratory of Electronics, M.I.T., October 15, 1963, p. 23.

2. M. Born and K. Huang, Dynamical Theory of Crystal Lattices (Oxford University Press, London, 1962), p. 82 ff.

3. A. S. Barker and J. J. Hopfield, Phys. Rev. 135A, 1732 (1964).

4. There is an error in the dispersion formula derived by Barker and Hopfield. They obtained $+2is_1 s_2 \omega \gamma_{12}$, instead of $-2is_1 s_2 \omega \gamma_{12}$, and had then to assume one of the s_ν negative for a reasonable fit of the experimental data.

5. H. Bilz, L. Genzel, and H. Happ, Z. Physik 160, 555 (1960); H. Bilz and L Genzel, Z. Physik 169, 53 (1962).

6. R. Wehner, Ph.D. Thesis, Freiburg, Germany, 1964.

V. ULTRASONIC PROPERTIES OF SOLIDS

Prof. C. W. Garland
R. H. Renard
C. F. Yarnell

A. ORDER-DISORDER LAMBDA TRANSITION IN AMMONIUM CHLORIDE AT HIGH PRESSURES

The lambda transition in NH_4Cl, which involves an ordering of the orientations of the NH_4^+ ion, has been extensively studied by the ultrasonic pulse technique. Both longitudinal and tranverse (shear) acoustic velocities have been measured over a wide range of frequencies (5-60 Mc) and temperatures (150-320°K) at 1 atm, with special emphasis on the region close to the lambda point (243°K). At a fixed frequency of 20 Mc and at five temperatures spaced between 250°K and 310°K, the pressure dependence of the velocities has now been measured between 1 atm and 12,000 atm, with emphasis on the region where the isotherms cross the lambda line. (At 310°K the lambda transition occurs at ~10,000 atm; thus our pressure work gives data on both the ordered and disordered phases at a series of temperatures.)

The results of these velocity measurements as a function of pressure are shown in Figs. V-1, V-2, and V-3. The McSkimin pulse-superposition method was used, and the values of the three independent elastic constants are known within an error of less than 0.05%. As seen in Figs. V-1 and V-2, the shear constants c_{44} and C' undergo a rapid increase when the pressure is increased near the lambda line. This variation is related to the volume changes that occur in the crystal upon ordering. Indeed, it can be shown that c_{44} and C' show very little variation with temperature or the degree of ordering if the volume is maintained constant. The anomalous variations of the shear constants with temperature, which were observed previously at constant pressure,[1] can now be shown to be due to the anomalous volume changes and not to the ordering per se. Since these shear constants provide a sensitive probe for the behavior of the volume upon ordering, it is interesting to note the rapidly changing character of the anomaly near the lambda point at high pressures.

Longitudinal waves are strongly coupled to the ordering process, and c_{11} values at 1 atm show an abrupt dip to a finite minimum at T_λ (and thus the adiabatic compressibility goes through a sharp but finite maximum).[1] Our new data in Fig. V-3 show this type of finite dip in c_{11} upon crossing the lambda line with increasing pressure. Note again the change in the behavior of the anomaly at higher pressures. For c_{11}, there is still an appreciable influence of the volume, but the entire anomaly is by no means all due to volume changes. Figure V-4 shows the variation of c_{11} with temperature at a constant pressure of 1 atm and at a constant volume of 34.15 cm^3 $mole^{-1}$. The most important difference between these two curves is the way in which the minimum value is approached from temperatures above and below. Thus, these new data, together with
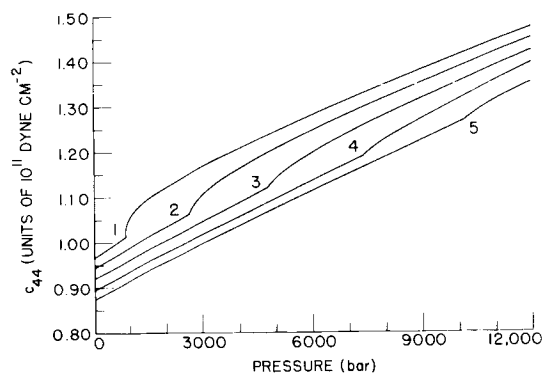
Fig. V-1. Shear elastic constant c_{44} versus pressure at (1) 250.7°, (2) 265.0°, (3) 280.0°, (4) 295.0°, and (5) 308.0°K.
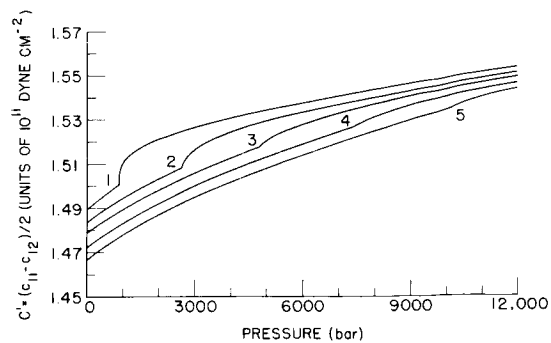

Fig. V-2. Shear elastic constant C' versus pressure at (1) 250.7°, (2) 265.0°, (3) 280.0°, (4) 295.0°, and (5) 308.0°K.
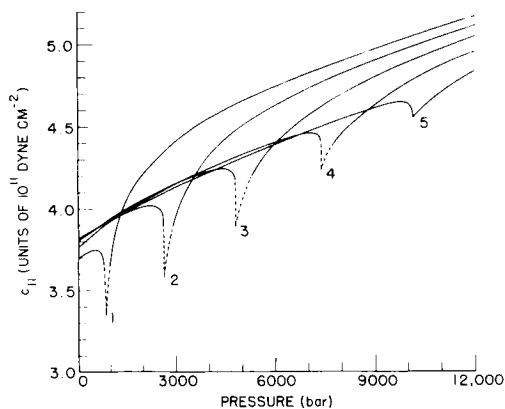

Fig. V-3. Compressional elastic constant c_{11} versus pressure at (1) 250.7°, (2) 265.0°, (3) 280.0°, (4) 295.0°, and (5) 308°K. Because of very high attenuation near the lambda line, data could not be obtained through the lambda point except at 308°K; for other temperatures the anticipated behavior is shown by the dashed lines.
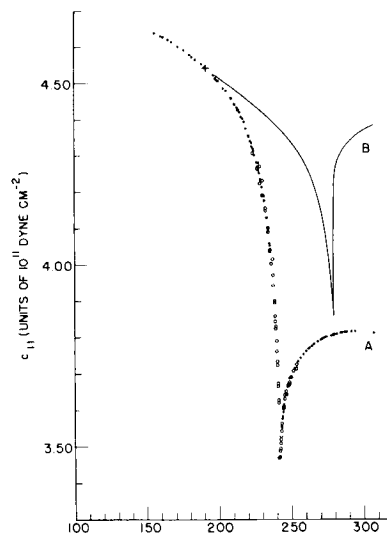

Fig. V-4. Variation of c_{11} with temperature at (a) constant pressure of 1 atm, and (b) constant volume of 34.15 cm^3 mole^{-1}. Solid dots represent new 1-atm measurements at 20 Mc/sec, and were obtained with the McSkimin method; open circles represent corrected data of Garland and Jones[1] at 5 and 15 Mc/sec.

our previous data, permit us to separate the effect of ordering, volume changes, and temperature changes. A detailed analysis of the effect of ordering alone (at constant volume and temperature) is in progress and will be reported elsewhere.

R. H. Renard

References

1. C. W. Garland and J. S. Jones, J. Chem. Phys. <u>39</u>, 2874 (1963).

VI. GEOPHYSICS

Prof. F. Bitter
Prof. G. Fiocco
Dr. T. Fohl
Dr. J. F. Waymouth

R. J. Breeding
J. C. Chapman
A. J. Cohen
J. B. DeWolf
R. J. DiGrazia

G. W. Grams
W. D. Halverson
H. C. Koons
K. Urbanek

A. PLASMA DIFFUSION IN A MAGNETIC FIELD

Complete results of experiments[1] on the diffusion of plasma of low-pressure arc discharges across a homogeneous magnetic field are now available.[2]

Further experiments are planned. Argon and Helium-Mercury discharges in homogeneous longitudinal fields to 25 kilogauss will be studied to obtain data on turbulent plasmas. Initial measurements will include determination of axial electric field and radial potential drop as a function of magnetic field. Spectral analysis and correlations of the RF noise and light fluctuations of the turbulent discharge will be performed by later experiments.

W. D. Halverson

References

1. W. D. Halverson, Plasma diffusion in a magnetic field, Quarterly Progress Report No. 75, Research Laboratory of Electronics, M.I.T., October 15, 1964, p. 31.

2. W. D. Halverson, Study of Plasma Columns in a Longitudinal Magnetic Field, Sc.D. Thesis, Department of Geology and Geophysics, M.I.T., January 1965.

B. OBSERVATIONS OF THE UPPER ATMOSPHERE BY OPTICAL RADAR IN ALASKA AND SWEDEN DURING THE SUMMER 1964 (Part II)

By invitation of the Geophysical Institute of the University of Alaska, and the Institute of Meteorology of the University of Stockholm, two optical radar devices were taken to Alaska and Sweden during the summer of 1964 to perform observations of the upper atmosphere at latitudes where noctilucent clouds might be visible. Some results were reported in Quarterly Progress Report No. 76 (pages 43-46).

Because optical radar observations require a dark sky background in order to maximize the signal-to-noise ratio, the unit in Sweden was mobile so that observations could be started at lower latitudes where there are longer periods of darkness. It was also of special interest to make observations during those nights when rocket experiments were being carried out at Kronogard, Sweden. These experiments, designed to measure temperature and wind velocity at the mesopause and to sample the particulate content of the noctilucent clouds, were carried out jointly by scientists of the Swedish Space Committee, the National Aeronautics and Space Administration, and the U.S. Air Force Cambridge Research Laboratories.

Visibility conditions were less than satisfactory much of the time in Alaska, as well as in Sweden. Precipitation was not uncommon and cirrus clouds or lower clouds were often present. On July 15, 1964, in Sweden, a very strong noctilucent cloud display was visible overhead and to the South, but the apparatus was not operational. On some other occasions, noctilucent clouds were seen low on the horizon, that is, at a distance of several hundred kilometers. All of our measurements were obtained when noctilucent clouds were not visible overhead. At our locations, during that period of time, we visually observed noctilucent clouds on the following nights:

5-6 August, Torsta

6-7 August, Torsta (strong display)

7 August, College

7-8 August, Torsta (strong display)

9-10 August, Torsta

10-11 August, Torsta

15-16 August, Torsta

16-17 August, Torsta

It is possible that on occasions noctilucent clouds may have been present, but were not observable from our location because of local visibility conditions.

The intensity of the echoes for a thin layer of homogeneously distributed scatterers, expressed as n_R, the expected number of photoelectrons per transmitted pulse emitted at the photodetector cathode, can be computed with the aid of the following expression:

$$n_R = \frac{\eta \lambda}{hc} W_t A_r k_r k_t k_a^2 \frac{HN\sigma}{4\pi R^2},$$

where W_t is the transmitted energy per pulse, A_r is the collecting area of the receiving telescope, k_r and k_t are the efficiencies of the receiving and transmitting systems, k_a is the atmospheric transmission, H is the geometrical depth of the layer, $N\sigma$ is the average radar cross section of the scatterers per unit volume, R is the distance, λ is the wavelength, η is the quantum efficiency of the photocathode, h is Planck's constant, and c is the speed of light. The product $HN\sigma$ is defined as the optical thickness of the layer. Because of the difficulty in measuring some of the parameters in the equation, we have calibrated the sensitivity of the apparatus by assuming that the echoes from heights of 40-60 km were produced by a model molecular atmosphere. This calibration was used to estimate the optical thickness of the layers above 60 km. It should be pointed out that the assumption of a dust-free atmosphere in the 40-60 km interval may lead to an underestimate of the optical thickness of these higher layers. On the other hand, our photoelectron counting procedure may be subject to a reduced counting efficiency at lower

heights, which would result in an overestimate of the optical thickness of the upper layers. While these and other systematic instrumental uncertainties may modify the results to a certain extent, the system does provide us with the capability of obtaining a quantitative measure of the optical thickness of scattering layers in the upper atmosphere.

The important source of noise is the sky background which varies throughout the night and from night to night. In order to illustrate the nature of this variability, Fig. VI-1 has been constructed to show, for a typical night, the solar depression angle and the intensity of the sky background as a function of time. The intensity of the background is displayed as the average number of photoelectrons emitted in a 66-μsec interval (equivalent to a resolution in distance of 10 km). The scatter in the points is due presumably to the occasional presence of thin cirrus clouds that increase the brightness of the background. Since the noise background increases very rapidly when the solar depression angle is less than 7°-8°, a very short interval of time is available for observations at high latitudes.
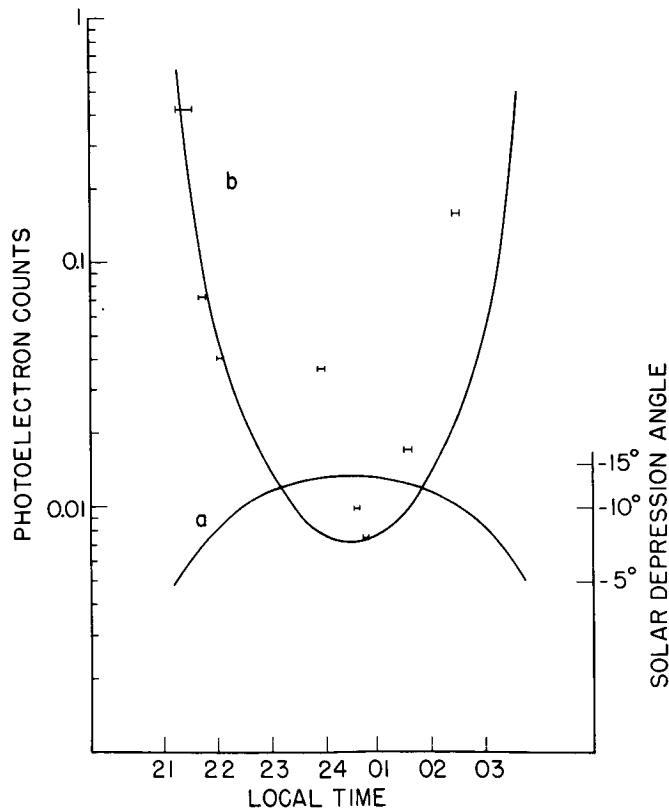

Fig. VI-1. Time variation of (a) solar depression angle, and (b) intensity of sky background for 16-17 August 1964 at Torsta, Sweden.
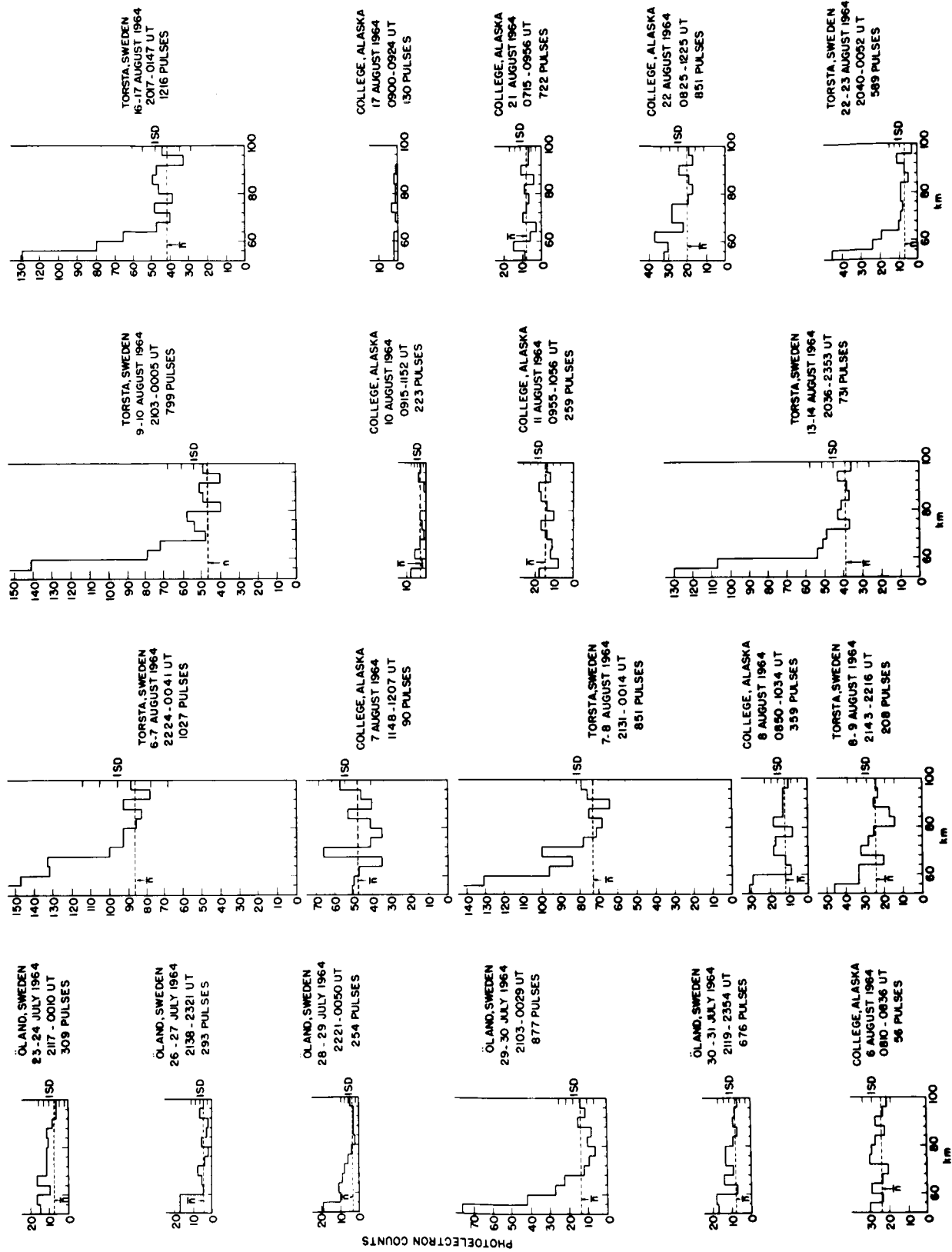
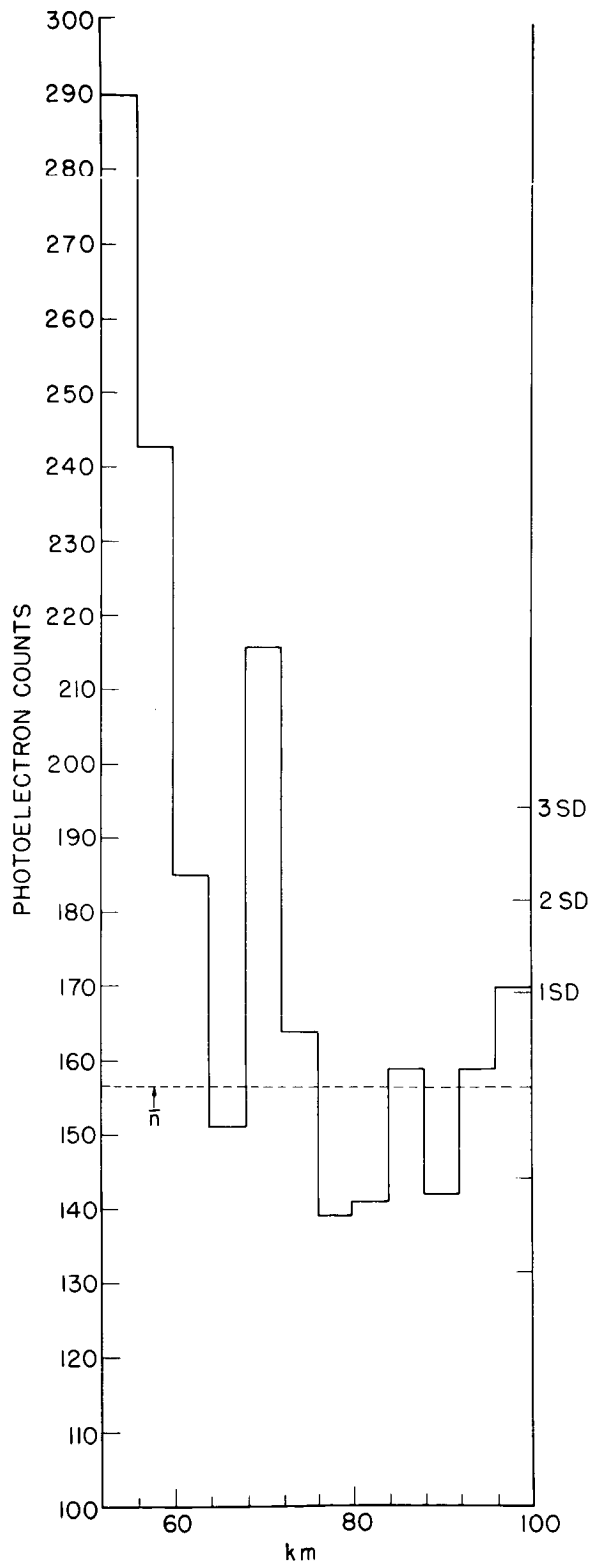Fig. VI-2. Summary of observation for summer 1964.

PHOTOELECTRON COUNTS

300
290
280
270
260
250
240
230
220
210
200
190 — 3 SD
180 — 2 SD
170 — 1 SD
160
150
140
130
120
110
100

$\bar{n}$

60   80   100

km

Fig. VI-3. Summary of observations obtained on 7, 7-8, 8, and 8-9 August 1964.

Figure VI-2 is a summary of all observations performed. For each night, the total number of photoelectrons accumulated in intervals of 4-km width from 52 km to 100 km is indicated, as well as the total number of pulses radiated and the Universal Time when the observations were made. The average noise level, $\bar{n}$, is taken as the average of the measurements between 88 km and 100 km. A measure of the statistical fluctuations, $\sqrt{n}$, is also indicated and is taken to be equivalent to the standard deviation of the measurement. Although we have a great number of traces with records from 100 km to 200 km, these will not be presented at this time. We point out, however, that these records do not show echoes of intensity comparable to those obtained in the summer of 1963 at those heights. The data presented in Fig. VI-2 indicate the presence of relatively strong echoes from an altitude of 68-72 km at both the Swedish and Alaskan locations on the nights of 7, 7-8, 8, 8-9 August, 1964. Note that we visually observed noctilucent clouds at both locations on the nights of 7 and 7-8 August. In order to gain statistical evidence that a scattering layer was present at that altitude, the observations for those four consecutive nights were averaged and the optical cross sections computed. The result is presented in Fig. VI-3. The optical thickness of this layer is estimated at $10^{-5} \pm 2 \times 10^{-6}$. Some measurements were also obtained on the night of 16-17 August in Torsta during a noctilucent cloud display. The partial count obtained from 0057 to 0147 U.T. has already been reported.[1] The data obtained during this time interval suggest the presence of scattering layers at altitudes of 60-64 km and 80-88 km. No comparable echoes were received during the earlier part of the night. This thickness of these layers is estimated to be 2×10^{-5} with a standard deviation of 1×10^{-5} for the 60-64 km layer, and 4×10^{-5} with a standard deviation of 2×10^{-5} for the 80-88 km layer.

Although these measurements are admittedly of a preliminary type, they indicate a value of the optical thickness for the noctilucent clouds substantially larger than that reported by Ludlam[2] and closer to the value suggested by Deirmendijan.[3]

G. Fiocco, G. W. Grams, K. Urbanek, R. J. Breeding

References

1. G. Fiocco, G. Grams, K. Urbanek, and R. J. Breeding, Quarterly Progress Report No. 77, Research Laboratory of Electronics, M. I. T., January 15, 1965, p. 45 (Fig. VI-3).

2. F. H. Ludlam, Noctilucent clouds, Tellus 9, 341-364 (1957).

3. D. Deirmendijan, Detection of Mesospheric Clouds from a Satellite, Proc. First International Symposium on Rocket and Satellite Meteorology, edited by H. Wexler and J. E. Caskey, Jr. (North-Holland Publishing Co., Amsterdam), pp. 406-411.

VII. GRAVITATION RESEARCH

Prof. R. Weiss
E. P. Jacobs
R. A. Sramek

RESEARCH OBJECTIVES AND PROJECT STATUS

Research in this group is concerned with investigating the nature of the gravitational interaction, in particular, with performing experiments that might distinguish between different theories of gravitation and cosmology. These aims are admittedly grandiose, especially in view of the inadequacy of present techniques which preclude performing many of the experiments that one can think of. The subject is, however, of such fundamental importance that even marginal experiments appear to be justified.

Our present intention is to perform an experiment to determine the constancy of G, the Newtonian gravitational constant. Interest in this has been motivated by a conjecture of Dirac[1] that G may have a secular variation of ~10^{-10} year because of the expansion of the universe. Jordan[2] and Dicke[3] have formulated scalar theories of gravitation within the framework of general relativity which have been tailored to fit Dirac's conjecture. Dicke's theory not only predicts a secular change in G but also an annual periodic variation of ~3×10^{-10} because of the Earth's eccentric motion around the sun.

Table VII-1. Limits on G variations.

Gravitational Force versus Inertial Reaction	$\frac{\Delta G}{G}$/year
Moon period versus earth rotation period coupled with theory of tidal torques for 200 years of observations and a check point against ancient eclipse data[4]	10^{-9}
Pendulum and Absolute Measurements of g made at Potsdam over 50 years[5]	10^{-7}
Gravitational Force versus Electric Force	
Spring gravimeters	10^{-7}
Cavendish experiment[6]	10^{-3}

Table VII-1 lists the limits that can be put on G variations from present knowledge. A distinction is made between limits set by measurements of a gravitational force against an inertial reaction and the gravitational force against another force field, as, for example, electromagnetic forces. It is conceivable that G variations would not be

observed in the first class of observations. The table is not exhaustive, studies of G variations as they might affect stellar evolution and the chronology of the Earth's history made by Dicke[7] and Jordan[8] do not offer smaller limits than those given in Table VII-1.

Proposed Experiment

The experiment that we propose is to measure variations in g, the Earth's gravitational attraction at the surface, with a stable gravimeter of a new design. Although the proposed method has the advantage that it measures a gravitational force against an electric force, it introduces the serious problem of the Earth's stability which will be discussed presently. All previous gravimeters employ Hooke's law forces in solids or gas pressure to balance the gravitational force on a mass. Measurements of g made with gas gravimeters fall in the first class, and are extremely temperature-sensitive. Spring gravimeters are troubled by temperature dependence of the Young's modulus of the spring material (the best give a temperature dependence $\frac{\Delta g}{g} \sim 10^{-6}/°C$), and in a much more sinister manner by random material creeps of the springs ($\Delta g/g \sim 10^{-8}$ month). In the proposed gravimeter (shown in Fig. VII-1) the spring is replaced by a measurable electric force and is fundamentally an inverted Kelvin absolute electrometer. Plate 1 is the
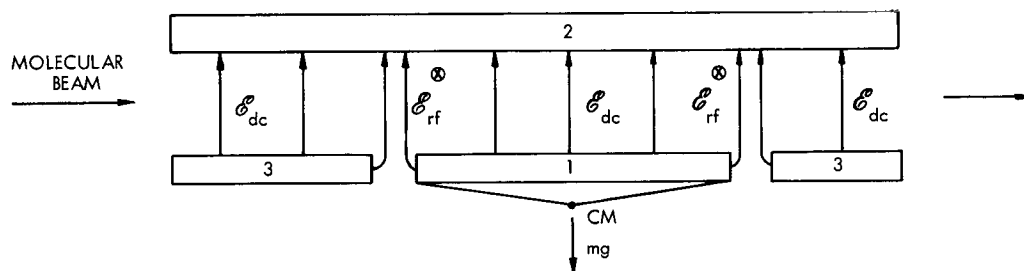

Fig. VII-1. Proposed gravimeter.

gravitating mass, which is supported vertically by the electric field $\mathscr{E}_{DC}$ and maintained horizontally by the fringing field between the guard plate 3 and the gravitating plate. The position of plate 1 is established interferometrically; the interferometer is incorporated in a null-seeking servomechanism which controls $\mathscr{E}_{DC}$ to maintain plate 1 co-planar with the guard plates. The servo maintains the relation

$$mg = F_e = \frac{1}{8\pi} \left\langle \mathscr{E}_{DC}^2 \right\rangle_p A,$$

where $\left\langle \mathscr{E}_{DC}^2 \right\rangle_p$ is the average of the square of the electric field over the surface, A,

of plate 1. In order to measure $\left\langle \mathscr{E}_{DC}^{-1} \right\rangle_p$ the gravimeter plates are incorporated in the transition region of a molecular beam electric resonance apparatus. $\Delta J = 0$, $Dm_J = \pm 1$ Stark transitions in a polar molecule are induced in separated regions of $\mathscr{E}_{RF}$ between plates 1 and 2. If the Stark energies are small compared with the rotational energy of the molecule, the Stark transition frequency is given by $\nu_{STARK} = k \left\langle \mathscr{E}_{DC}^2 \right\rangle_B$, where k is a function of the molecular constants and quantum numbers and $\left\langle \mathscr{E}^2 \right\rangle_B$ is the average of the square of the electric field at the location of the molecular beam between the regions of $\mathscr{E}_{RF}$. Assuming that $\left\langle \mathscr{E}_{PL}^2 \right\rangle_B = \left\langle \mathscr{E}_{DC}^2 \right\rangle_p$ (which is not strictly so because of the fringing field and, for only one molecular beam, because of nonparallelism of the plates) we can save the force equation and get

$$\nu_{STARK} = \frac{8\,k}{A}\,mg.$$

If m and A are constant, the g measurement becomes a frequency measurement

$$\frac{\Delta g}{g} = \frac{\Delta \nu}{\nu}.$$

This rudimentary description of the gravimeter only alludes to a host of problems which must be investigated first. Among these are:

1. Constancy of the gravitating mass and temperature dependence of the plate area.

2. Constancy of the relation between F_e and $\left\langle \mathscr{E}^2 \right\rangle_B$. In particular, how does the relation vary with plate separation, tilt, and co-planarity of the gravitating plate and the guard plates? An experiment is now in progress to study these effects for various plate geometries and potential distributions.

3. Field homogeneity conditions necessary to see a resonance. Inhomogeneitics can arise from nonparallelism of the plates and the time dependence of ground noise. For example, it may be necessary to have a fast AC servo to hold the plate, and hence $\mathscr{E}_{DC}^2$, steady over the integration time of the servo that scans the resonance.

Stability of the Earth

Although no measurements have been reported of changes in g at a fixed point on the Earth (besides calculable and known effects such as the tides), it is likely that they exist, especially within the precision demanded for this experiment. In principle, the only multipole moment of the Earth's gravitational potential whose changes are not amenable to averaging over distributed gravimeter sites is the monopole. A catalogue of the mass changes that one can think of would yield as the largest contribution $\Delta g/g \sim 10^{-16}$/year from meteor influx. Estimates of changes in the Earth's radius are far more influential. Again, there are no measurements, and some theories predict expansion, and others, contraction. Jeffreys,[9] and MacDonald[10] are exponents of

contraction caused by cooling which would result in changes in radius producing $\frac{\Delta g}{g} \sim +10^{-10} \rightarrow 10^{13}$ per year. Egyed[11] calculates for an expansion, together with continental drift, a change in radius that could result in a $\frac{\Delta g}{g} \sim -10^{-10}$ per year. To emphasize this sensitivity, one need only observe that a change of 0.3 mm in the radial position of the gravimeter corresponds to a $\frac{\Delta g}{g} \sim 10^{-10}$.

One change in the quadrupole moment can already be anticipated. The melting of the polar ice caps results in a $\frac{\Delta g}{g} \sim 10^{-10}$ per year at 45° latitude; however, it is zero at 30° latitude. Variations in the Earth's rotation rate are monitored to the precision necessary for the experiment.

Besides the Earth's radius, the other major unknowns are the magnitudes of local distortions and mass redistributions in the vicinity of a given site (the very high order multipole moments). Knowledge of this will set a limit on the number of sites necessary to establish a statistical estimate of the global g variation. We would endeavor, of course, to pick seismically inactive rock shields with small gravity anomalies. To our knowledge, the only measurements of secular earth distortion (except in regions of post-glacial uplift) have been made by the Benioff Strain seismometer at the Lamont Geophysical Laboratory site, in Ogdensberg, New Jersey, where a limit of $\frac{\Delta \ell}{\ell} < 10^{-8}$/month over a 200-ft length[12] has been set. It appears, however, that the limit is instrumental.

In order to have a handle on radius changes and to correlate local distortions with changes in g, it would be useful to have a stable strain seismometer associated with the gravimeter. We are, at present, engaged in the design and construction of a stabilized laser strain seismometer. Another necessary instrument is a stable tiltmeter, since the gravimeter is quadratically dependent on the angle between the plumb and the normal to the gravitating plate. This instrument is being constructed as part of senior thesis research.

Our first effort will be to build one gravimeter and search for the annual periodic change in g predicted by Dicke's scalar theory.

R. Weiss

References

1. P. A. M. Dirac, Proc. Roy. Soc. (London) A165, 199 (1938).

2. P. Jordan, Schwerkraft und Weltall (Friedrich Vieweg und Sohn, Braunschweig, 1955).

3. C. Grans and R. H. Dicke, Phys. Rev. 124, 1925 (1961).

4. W. H. Munk and G. J. F. MacDonald, The Rotation of the Earth (Cambridge University Press, London, 1960).

5. G. P. Woollard, Advances in Geophysics, Vol. 1, p. 281, 1952.

6. F. H. Newman and V. H. L. Searle, The General Properties of Matter (Edward Arnold Ltd., London, 5th edition, 1957).

7. R. H. Dicke, Rev. Modern Phys. 34, 110 (1962).

8. P. Jordan, Naturwiss. <u>48</u>, 417 (1961).

9. H. Jeffreys, <u>The</u> <u>Earth</u> (Cambridge University Press, London, 4th edition, 1962).

10. G. J. F. MacDonald, Revs. Geophys. <u>1</u>, 587 (1963).

11. L. Egyed, Nature <u>178</u>, 534 (1956).

12. M. W. Major, G. H. Sutton, U. Oliver, and R. Metsger, Bull. Seismol. Soc. Am. <u>54</u>, 295 (1964).

VIII. MAGNETIC RESONANCE

Prof. J. S. Waugh
Dr. J. Rugheimer
Dr. R. Newmark
J. D. Macomber

J. W. Riehl
C. G. Wade
D. W. Schaefer

D. S. Thompson
E. T. Stone
A. Leonardi
E. L. Wei

A. SELF-DIFFUSION IN LIQUID ETHANE

Measurements have been completed of the self-diffusion coefficient D of liquid C_2H_6 at temperatures between 155° and 298°K and between 1 and 2500 atm (Fig. VIII-1).
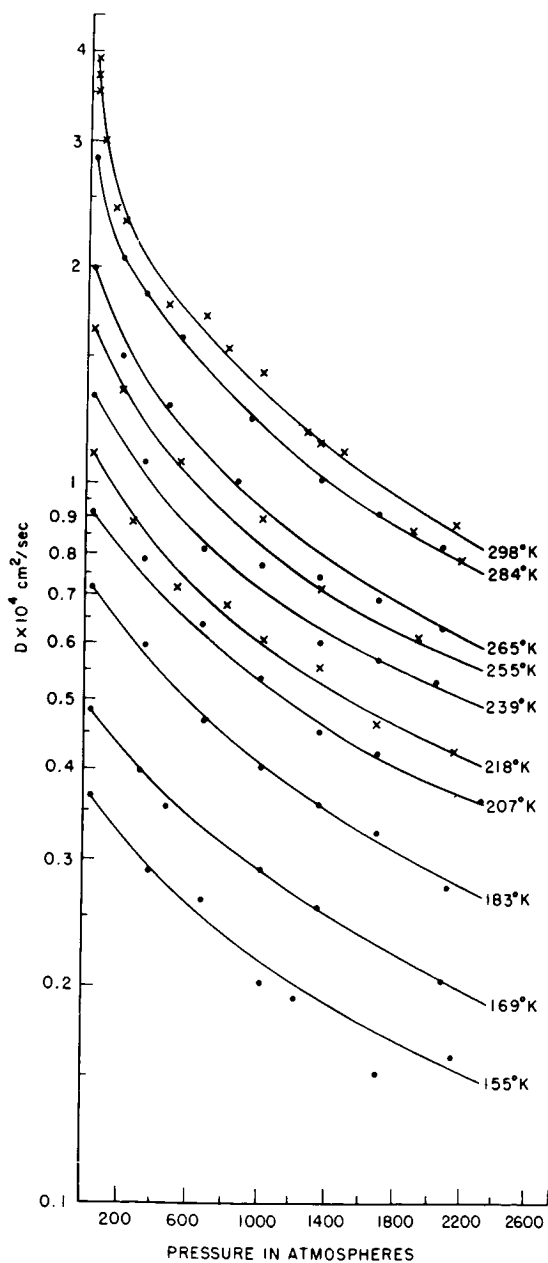

Fig. VIII-1. Dependence of the self-diffusion coefficient D in ethane on pressure at various constant temperatures.

Proton spin echoes were used for this purpose.[1] The equation of state was determined very approximately by measuring the pressure dependence of free-induction decay amplitudes at various constant temperatures. Thus it was possible to show (as anticipated) that the free-volume theory, which could be used to fit our earlier results at constant pressure,[2] is inconsistent with the experimental density dependence. Figure VIII-2 shows
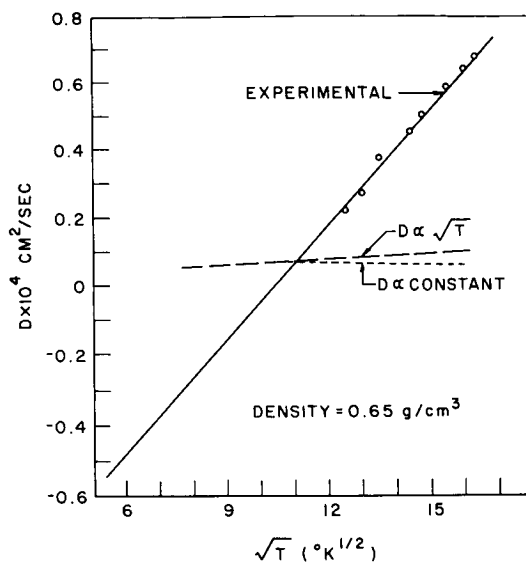

Fig. VIII-2. Experimental and theoretical values of D at constant volume.

representative results reduced to constant density. The circles are experimental points. The dotted and dashed lines show the predictions of the Doolittle[3] and of the Cohen-Turnbull[4] versions of free-volume theory, with the parameters that fit our previous measurements at constant pressure[2] used.

C. G. Wade, J. S. Waugh

References

1. E. L. Hahn, Phys. Rev. 80, 580 (1950).
2. J. V. Gaven, W. H. Stockmayer, and J. S. Waugh, J. Chem. Phys. 37, 1188 (1962).
3. A. K. Doolittle, J. Appl. Phys. 22, 1471 (1951).
4. M. H. Cohen and D. Turnbull, J. Chem. Phys. 31, 1164 (1959).

IX. PHYSICAL ACOUSTICS*

Prof. K. U. Ingard
Prof. L. W. Dean III
Dr. G. C. Maling, Jr.
Dr. H. L. Willke, Jr.

P. A. Fleury V
K. W. Gentle
J. L. Macon
A. A. Maduemezia
W. M. Manheimer

M. A. Martinelli
J. A. Ross
T. B. Smith
S. D. Weiner

A. RESPONSE OF LEAD ZIRCONATE TITANATE CERAMICS TO TEMPERATURE FLUCTUATIONS

1. Introduction

It has been found that some shapes of lead zirconate titanate (PZT) ceramics respond to rapid temperature fluctuations when mounted in a particular geometrical configuration. This effect is presumably caused by thermal stresses in the crystal which produce a voltage difference across the silvered surfaces.

These units were originally constructed to serve as pressure transducers in a low-pressure shock tube, and although a response was obtained when a shock wave (Mach number $\doteq$ 1.25) passed the transducer, the waveform was quite different from the waveform produced by a commercial pressure transducer. Investigation into the differences between the waveforms led to the conclusion that the transducer was sensitive to the temperature rise behind the shock wave.

We report some of the results obtained with the transducer in the shock tube, describe a calibration made with a hot-air jet and a mechanical chopper wheel, and show some sample waveforms produced by spoken digits.

2. Description of the Transducers

Three particular configurations have been tested. The first transducer constructed is shown in Fig. IX-1c. It consists of a PZT tube mounted on the end of a section of hypodermic needle. The tube dimensions are 1/8 inch in diameter and 1/8 inch long. The outside of the hypodermic tubing and the inside of the ceramic tube are connected by conducting epoxy adhesive and form the ground lead. The second lead is passed through the center of the hypodermic tubing, and is attached (with conducting epoxy) to the outside of the ceramic tube. A connector is attached to the other end of the hypo-dermic tube. All of the results reported here have been obtained with this transducer.

The transducer shown in Fig. IX-1b has been found to have essentially the same properties as the transducer described above; no detailed measurements have been made. The construction is essentially similar except for the size and the fact that the conductor inside the PZT tube is much smaller than the inside diameter of the tube.
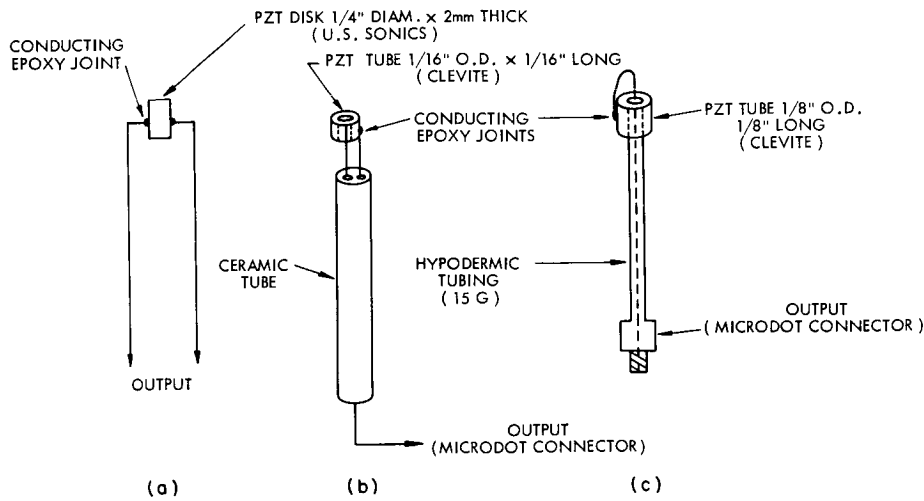
Fig. IX-1. Three types of thermal transducers.

A third unit, shown in Fig. IX-1a, has also been tested in the chopped air jet described below. We were not able to measure any response to temperature fluctuations with this unit.

3. Calibration with a Chopped Air Stream

The apparatus shown in Fig. IX-2 was constructed in order to measure the temperature response of these transducers.

The air jet is heated by the gas flame and is chopped by the wheel and 1600 rpm motor. The jet is interrupted approximately 100 times per second.
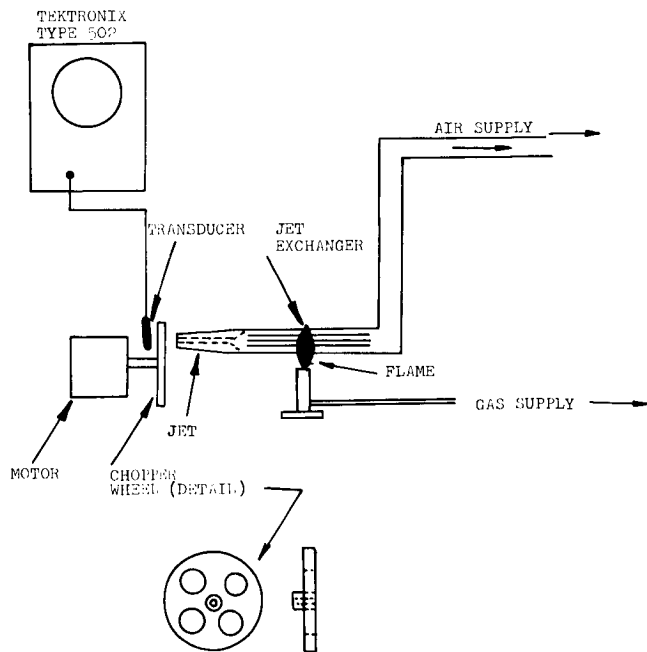

Fig. IX-2. Schematic diagram of the chopper apparatus.

Some experimental results are shown in Fig. IX-3. The air jet temperature is approximately 80°F, and the approximate velocities are marked on the figure. It was not possible to obtain a very accurate estimate of the flow velocity because of air leaks in the heat exchanger. The oscilloscope vertical sensitivity is given on the figure, and the sweep speed was 5 msec/cm. The smooth portions of the trace occur when the jet is off. When the jet is on the response is caused partly by the vibrations induced in the hypodermic tubing by the turbulent air stream. The maximum levels that occur for any stream velocity are approximately 10 millivolts.

(a) VERTICAL SENSITIVITY,
5 mv/cm;
JET VELOCITY,
18 m/sec.

(b) VERTICAL SENSITIVITY,
10 mv/cm;
JET VELOCITY,
30 m/sec.

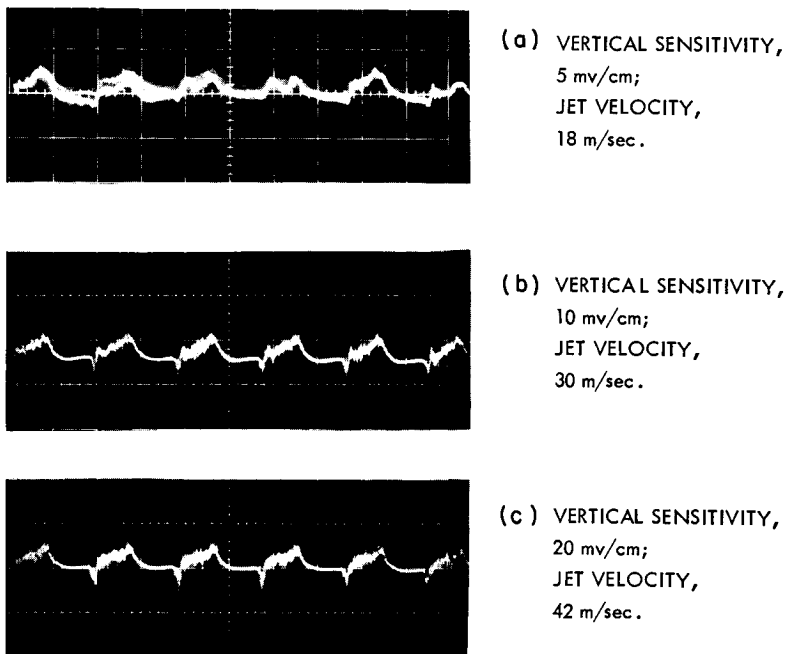(c) VERTICAL SENSITIVITY,
20 mv/cm;
JET VELOCITY,
42 m/sec.

Fig. IX-3. Transducer output produced by a cold jet.

When the jet is heated by the burner, much larger voltages are produced. Some examples are shown in Figs. IX-4 and IX-5. The top portions occur when the warm jet is cut off, and the bottom portions occur when the warm jet impinges on the transducer. In all cases, the oscilloscope vertical sensitivity is 20 mv/cm, and the sweep speed is 5 msec/cm. The temperature of the air stream is given for each photograph. A sensitivity curve was made from these data, and is shown in Fig. IX-6. It can be seen that the output voltage appears to be linear with temperature rise, except near the maximum levels that were measured. More data — particularly a more precise method of measuring the temperature — are needed in order to obtain a better determination of the sensitivity curve.
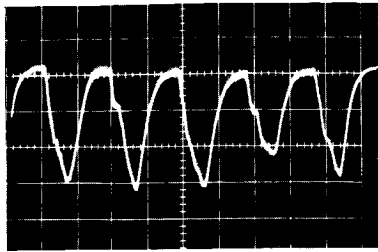
(a)   JET TEMPERATURE,
      = 160°F.
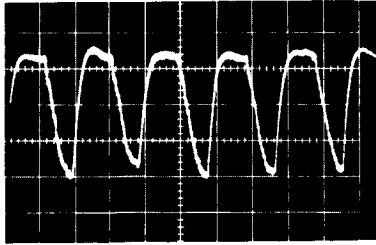
(b)   JET TEMPERATURE,
      = 145°F.

Fig. IX-4.
Response of the transducer to
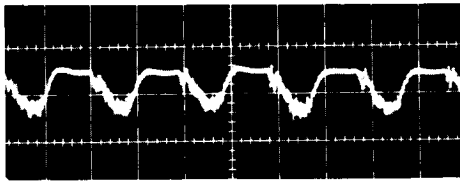a warm chopped air stream.
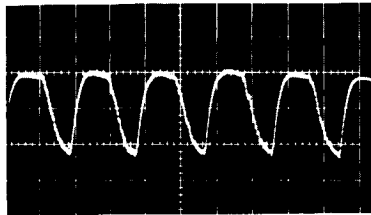
(c)   JET TEMPERATURE,
      = 130°F.

Fig. IX-5.
Response of the transducer to a
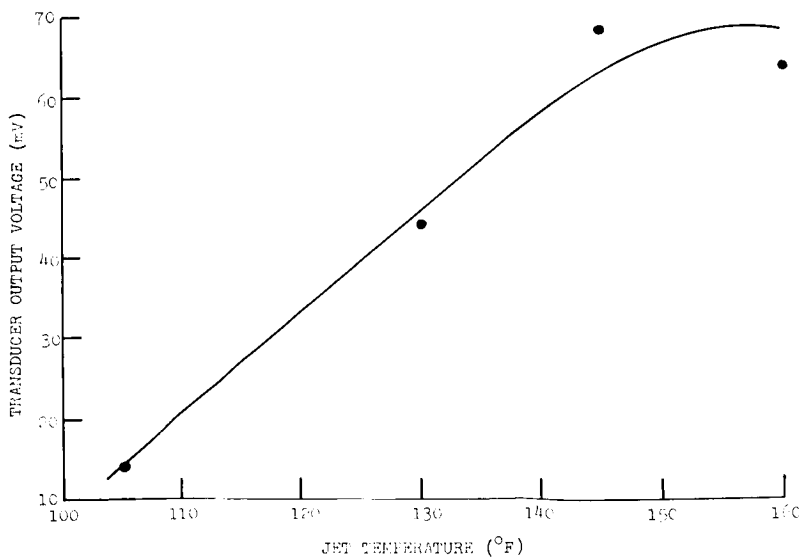warm chopped air stream. Jet
temperature, 105°F.

Fig. IX-6.
Plot of the measured temperature sensitivity of the temperature transducer. (Data from Figs. IX-4 and IX-5.)
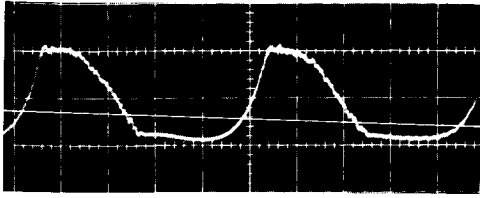
Fig. IX-7. Wave shape for finding the rise time of the thermal transducer.

Finally, the rise time of the transducer can be determined approximately from the waveform shown in Fig. IX-7. Here, the air jet temperature was 125°F, the sweep speed was 2 msec/cm, and the vertical sensitivity was 20 mv/cm. The rise time (the "fall time" since the trace polarity is inverted) is seen to be approximately 1.4 msec, which corresponds to an upper cutoff frequency of approximately 700 cps. It has been found in other experiments that the transducer does not respond to a steady temperature rise. No indication of the decay after a temperature rise can be seen from this figure, however, because the chopping speed is too fast.

4. Response to a Shock Wave

The transducer was placed in a shock tube in order to determine its response to a temperature (and pressure) rise. Figure IX-8 is a schematic diagram of the tube and the associated instrumentation. The pressure transducer is located in the end of the tube; the PZT transducer is located approximately 4 inches from it in the center of the tube. Both pressure and temperature waveforms were recorded simultaneously on the dual-beam oscilloscope. The results are shown in Figs. IX-9, IX-10, and IX-11. The lower trace is the output of the DYNAGAGE pressure transducer, and the upper trace
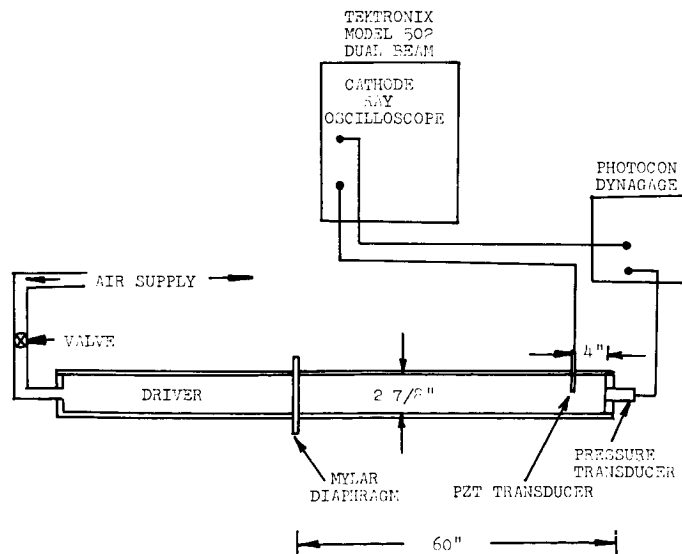

Fig. IX-8. Schematic diagram of the shock-tube apparatus.

is the PZT transducer output. The sensitivity S_p of the pressure transducer is given for each waveform (the units are volt/psi). From these data the amplitude of the pressure pulse can be found.
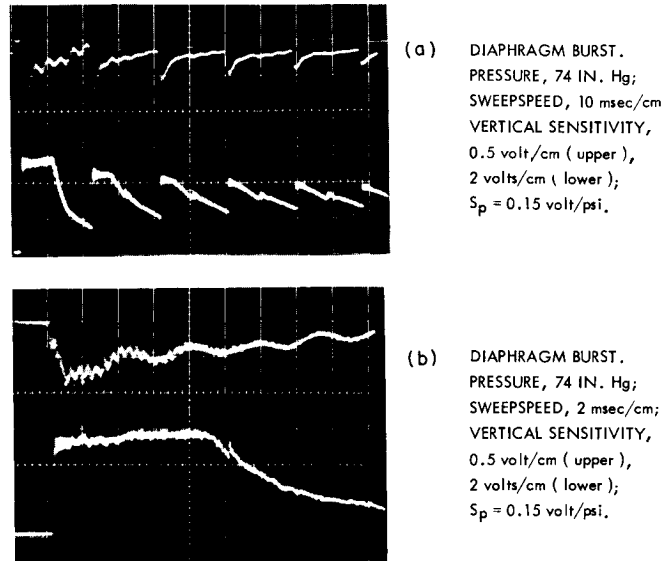

(a) DIAPHRAGM BURST.
PRESSURE, 74 IN. Hg;
SWEEPSPEED, 10 msec/cm
VERTICAL SENSITIVITY,
0.5 volt/cm (upper),
2 volts/cm (lower);
$S_p = 0.15$ volt/psi.

(b) DIAPHRAGM BURST.
PRESSURE, 74 IN. Hg;
SWEEPSPEED, 2 msec/cm;
VERTICAL SENSITIVITY,
0.5 volt/cm (upper),
2 volts/cm (lower);
$S_p = 0.15$ volt/psi.

Fig. IX-9. Response of the pressure transducer and PZT transducer to a shock wave generated by a mylar diaphragm that bursts at 74 in. Hg.

(a) DIAPHRAGM BURST.
PRESSURE, 47 IN. Hg;
SWEEPSPEED, 10 msec/cm;
VERTICAL SENSITIVITY,
0.2 volt/cm (upper),
1 volt/cm (lower);
$S_p = 0.15$ volt/psi.

(b) DIAPHRAGM BURST.
PRESSURE, 47 IN. Hg;
SWEEPSPEED, 2 msec/cm;
VERTICAL SENSITIVITY,
0.2 volt/cm (upper),
1 volt/cm (lower);
$S_p = 0.15$ volt/psi.

Fig. IX-10. Response of the pressure transducer and PZT transducer to a shock wave generated by a mylar diaphragm that bursts at 47 in. Hg.

(a) DIAPHRAGM BURST.
PRESSURE, 11 IN. Hg;
SWEEPSPEED, 10 msec/cm;
VERTICAL SENSITIVITY,
50 mv/cm (upper),
0.2 v/cm (lower),
S_p = 0.15 volt/psi.

(b) DIAPHRAGM BURST.
PRESSURE, 11 IN. Hg;
SWEEPSPEED, 2 msec/cm;
VERTICAL SENSITIVITY,
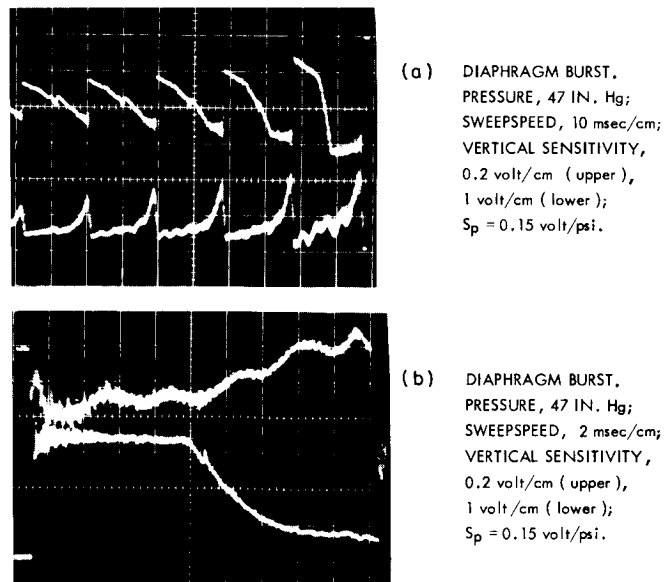50 mv/cm (upper),
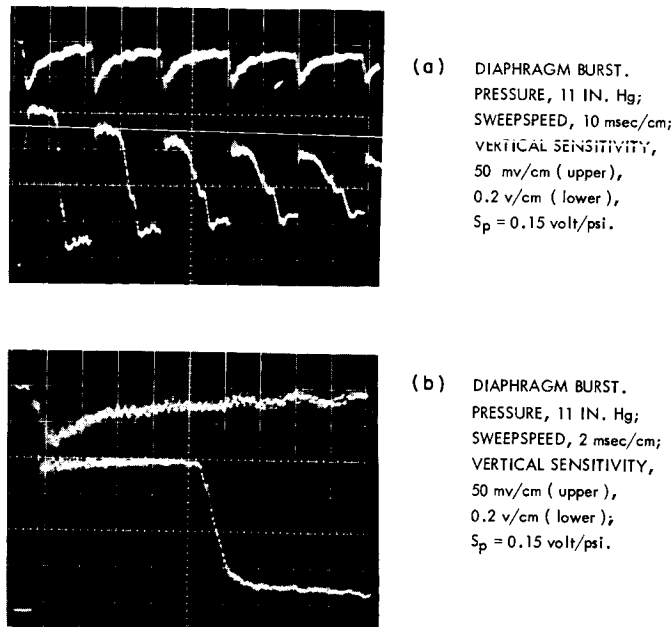0.2 v/cm (lower);
S_p = 0.15 volt/psi.

Fig. IX-11. Response of the pressure transducer and PZT transducer to a shock wave generated by a mylar diaphragm that bursts at 11 in. Hg.

Other time-of-flight measurements have been made[1] to determine the Mach number of the incident shock. From these data, the temperature behind the shock can be calculated.[2] The results are summarized in Table IX-1. These voltages are much larger

Table IX-1. Results of calculations.

Burst pressure (in. Hg)	Shock wave Mach number	Temperature rise (°F)	PZT transducer output (volts)
11	1.04	16	.055
46	1.10	38	.3
74	1.27	100	.75

than the voltages produced by the chopped air stream, even though the pressure rise is the same order of magnitude. The reason for this discrepancy is not understood.

Figures IX-9, IX-10, and IX-11 also show a voltage decay behind the temperature rise, and some oscillations. The voltage decay may be caused by a lack of low frequency response of the transducer, but the oscillations that occur during decay are not understood at this time.

5. Speech Waveforms

One interesting application of this ceramic transducer is to measure temperature fluctuations in front of a person's mouth as he speaks.
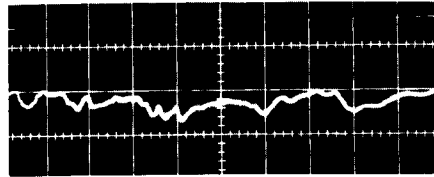
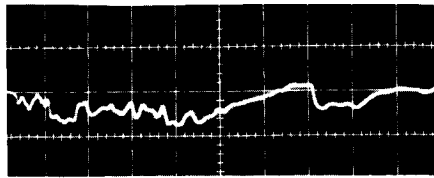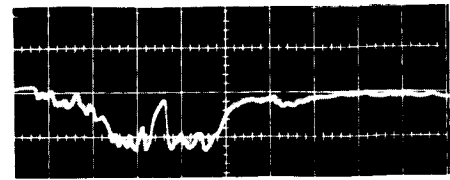Fig. IX-12. Waveform for digit zero.
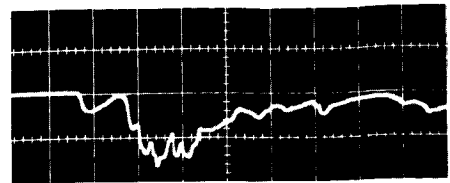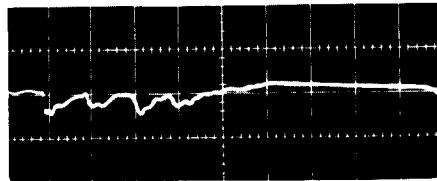

Fig. IX-13. Waveform for digit one.
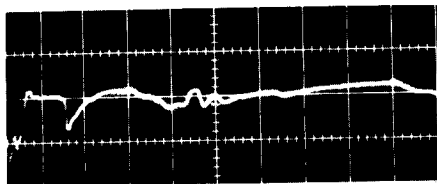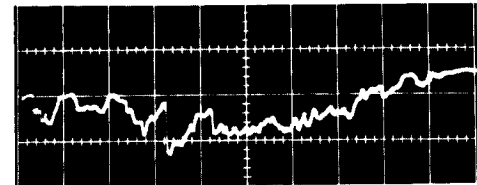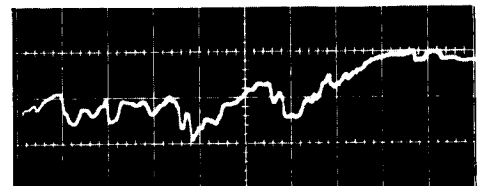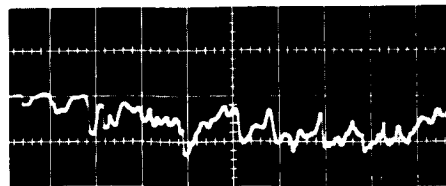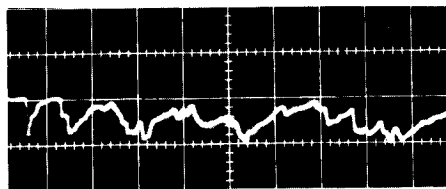

Fig. IX-14. Waveform for digit two.

Fig. IX-15. Waveform for digit three.
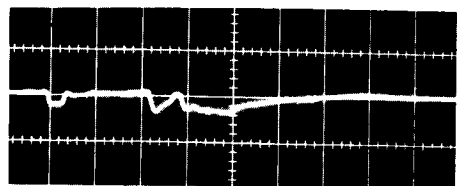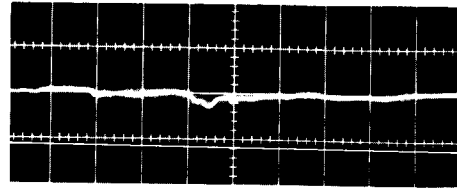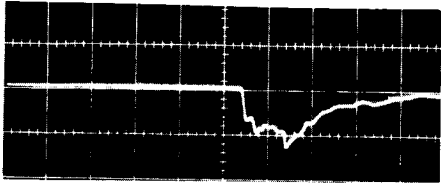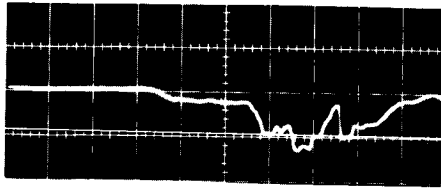

Fig. IX-16. Waveform for digit four.

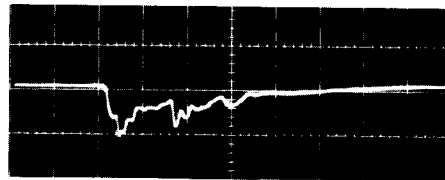Fig. IX-17. Waveform for digit five.

Fig. IX-18. Waveform for digit six.
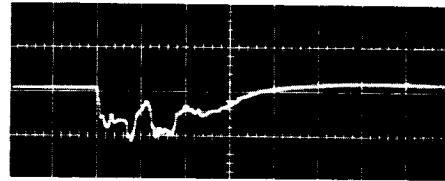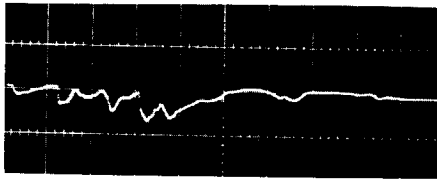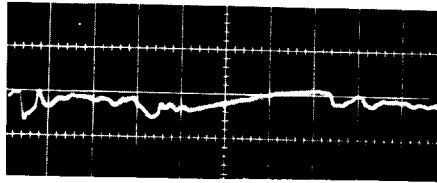

Fig. IX-19. Waveform for digit seven.

Fig. IX-20. Waveform for digit eight.
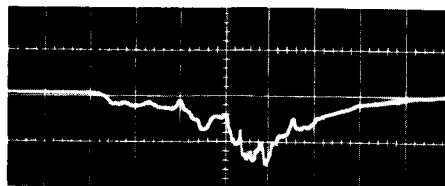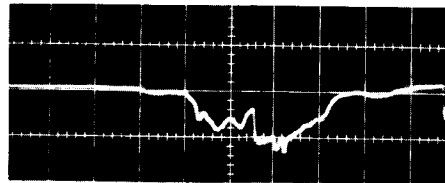
Fig. IX-21. Waveform for digit nine.

This application has not been investigated thoroughly, but some preliminary results for spoken digits zero through nine are shown in Figs. IX-12-IX-21. Two waveforms of each number are shown in order to estimate the repeatability of the waveform. It can be seen that not all of the features are repeatable, although the general shape of the waveform is the same. The PZT transducer was held in the hand when the photographs were taken, and this may account for some of the lack of repeatability.

The oscilloscope was triggered by a condenser microphone placed ~1 ft from the speaker's mouth. All data were obtained in the anechoic chamber of the Research Laboratory of Electronics.

In all of the figures, the sweep speed was 50 msec/cm, and the vertical sensitivity was 20 mv/cm.

Each digit appears to have a reasonably characteristic waveform, and the fluctuations take place rather slowly. Waveforms produced by vowels or other speech sounds have not been investigated.

G. C. Maling, Jr., U. Ingard

References

1. G. C. Maling, Jr. and U. Ingard, Shock wave transmission and attenuation, Quarterly Progress Report No. 75, Research Laboratory of Electronics, M.I.T., October 15, 1964, pp. 42-44.

2. G. Rudinger, Wave Diagrams for Non-Steady Flow in Ducts (D. Van Nostrand Company, Inc., New York, 1955).

B. SPONTANEOUS INSTABILITY IN PARALLEL FLOWS

Quarterly Progress Report No. 76, p. 75 contained a simple proof of the instability of all time-symmetric systems, including the inviscid, parallel flows. The argument used did not distinguish between "spontaneous" instability, that is, the amplification up to macroscopic levels of perturbations initially smaller than any given positive bound, and "passive" instability, that is, the permanent alteration of the character of the flow under gross perturbations. We shall exhibit a class of such "infinitesimal" perturbations for inviscid, stationary, parallel flows, or parallel plasmas with collinear magnetic fields, which always yield "spontaneous" instability. This work is a generalization of an earlier result (Quarterly Progress Report No. 68, p. 45) for incompressible flow.

We begin by demonstrating that stationary, nondissipative, parallel flows are in fact homogeneous along streamlines. For a velocity field $\underline{v}(\underline{x}) = \underline{\epsilon}_1 w(\underline{x})$, and if the fluid be a plasma, a magnetic field $\underline{B}(x) = \underline{\epsilon}_1 B_1(x_2, x_3)$, the conservation equation for mass becomes $\nabla \cdot \rho \underline{v} = \partial_1 \rho w = 0$, so that

$$\rho w(\underline{x}) = m(x_2, x_3). \tag{1}$$

From the momentum conservation equation, we have also

$$\rho w \partial_1 w + \partial_1 \left(p + \frac{B^2}{8\pi} \right) - \frac{1}{4\pi} B_1 \partial_1 B_1 = m \partial_1 \frac{m}{\rho} + \partial_1 p = \partial_1 \left(\frac{m^2}{\rho} + p \right) = 0$$

so that

$$p(\underline{x}) + \frac{m^2(x_2, x_3)}{\rho(\underline{x})} = f(x_2, x_3). \tag{2}$$

For a thermally nonconducting fluid, we have $\frac{dS}{dt} = w \partial_1 S = 0$, with the result that $S = S(x_2, x_3)$. If the equation of state

$$p(\underline{x}) = p(\rho(\underline{x}), S(x_2, x_3)) \tag{3}$$

for the fluid at $\underline{x}$ is independent of (2), then (2) and (3) together determine a set of isolated solutions for $p(\underline{x})$ and $\rho(\underline{x})$. Moreover, since (2) and (3) are independent of x_1, p and ρ will be also. Thus p, ρ, S, and $v = \frac{m}{\rho}$ are all constant on streamlines.

The exceptional case, for which (2) and (3) are, for some range of p and ρ, not independent, can occur only if

$$p(\rho, S) = a(S) - \frac{b(S)}{\rho} \tag{4}$$

over that range, and if in addition $f(x_2, x_3) = a(S(x_2, x_3))$ and $m^2(x_2, x_3) = b(S(x_2, x_3))$ for some S. The equation of state (4) cannot be excluded on thermodynamic grounds; however, it does contradict physical experience. The speed of sound for such a substance would be

$$c^2 = \left(\frac{\partial p}{\partial \rho} \right)_S = \frac{b(S)}{\rho^2}, \tag{5}$$

whereas for real substances the speed of sound increases with density. It is precisely this decrease with density which in fact enables stationary density and pressure variations to persist in spite of the convective action of the flow. For, combining (1), (5), and the second relation following (4), we get

$$c^2(\underline{x}) = \frac{b(S(x_2, x_3))}{\rho^2(\underline{x})} = \frac{m^2(x_2, x_3)}{\rho^2(\underline{x})} = v^2(\underline{x})$$

so that the flow velocity adjusts itself to bring an upstream acoustic wave to rest at each point.

This entire argument applies to perfectly heat-conducting fluids, the only remaining nondissipative type, if S is replaced by T, which is necessarily independent of the x_i.

We are now in a position to demonstrate the spontaneous instability of nondissipative, stationary, parallel flows, except for the peculiar, nonhomogeneous type discussed above; namely, to develop a class of infinitesimal perturbations that, in lowest order, grow without bound.

It is not hard to show that in order to yield such instability, a perturbation must involve the velocity field. For the present case we shall see that it suffices to perturb only the velocity field.

Given the parallel flow $(\underline{v} = \underline{\epsilon}_1 w, \rho, S)$, then $(\underline{v}+\underline{u}, \rho, S)$ will also be a flow provided that $\underline{u}$ satisfies in lowest order

$$\nabla \cdot \rho\underline{u} = 0, \tag{6a}$$

$$\frac{\partial\underline{u}}{\partial t} + (\underline{v} \cdot \nabla)\underline{u} + (\underline{u} \cdot \nabla)\underline{v} = 0, \text{ and} \tag{6b}$$

$$(\underline{u} \cdot \nabla)S = 0 \qquad \text{(thermally nonconducting fluid only)}. \tag{6c}$$

For unperturbed flows independent of x_1, Eqs. 6 become cyclic in x_1. It follows that $\underline{u}$ will always be independent of x_1 if it is initially. Taking $\underline{u} = \underline{u}(x_2, x_3)$, we get

$$\partial_2\rho u_2 + \partial_3\rho u_3 = 0 \tag{7a}$$

$$\frac{\partial u_1}{\partial t} + u_2\partial_2 w + u_3\partial_3 w = 0 \tag{7b}$$

$$\frac{\partial u_2}{\partial t} = 0 \tag{7c}$$

$$\frac{\partial u_3}{\partial t} = 0 \tag{7d}$$

$$(u_2\partial_2 S + u_3\partial_3 S = 0) \tag{7e}$$

According to (7c) and (7d), $u_2(\underline{x}, t) = u_2^{(0)}(x_2, x_3)$ and $u_3(\underline{x}, t) = u_3^{(0)}(x_2, x_3)$, so that Eq. 7a (and, if appropriate, Eq. 7e) will always be satisfied if it is initially. We can always satisfy (7a) and (7e) by taking $\rho u_2^{(0)} = \epsilon\partial_3 S$ and $\rho u_3^{(0)} = -\epsilon\partial_2 S$, i.e. by using in the plane a stream function proportional to the entropy. This measure is not necessary in isentropic regions or in perfectly heat-conducting fluids.

The remaining Eq. 7b integrates at once to

$$u_1(\underline{x}, t) = u_1^{(0)} - tu_2^{(0)}\partial_2 w - tu_3^{(0)}\partial_3 w. \tag{8}$$

We have arrived at what may be called a purely convective instability in nondissipative parallel flows, namely, a class of perturbations that are constant along streamlines, and involve only the velocity field, which, in lowest order, produce linearly

growing disturbances of the velocity field only. The streamlines $\underline{\epsilon}_1 w(x_2, x_3)$ of the unperturbed flow are carried bodily along the streamlines of the steady, perturbing, convective field (u_2, u_3), so that the change in u_1 at $\underline{x}$ is precisely the change in the strength w of the streamline of the unperturbed flow as it is convected past that point and, to lowest order, equals the gradient of w against the streamline of (u_2, u_3) multiplied by the displacement along the streamline or roughly $t \sqrt{u_2^2 + u_3^2}$, in agreement with (8).

It is illuminating in this connection to replace (7b) by the exact equation for $V_1 = w + u_1$,

$$\frac{\partial V_1}{\partial t} + u_2^{(0)} \partial_2 V_1 + u_3^{(0)} \partial_3 V_1 = 0 \tag{9}$$

which has the solution

$$V_1(x_2, x_3, t) = V_1^{(0)}(X_2, X_3), \tag{10}$$

where $\underline{X}$ is the Lagrangian coordinate associated with $\left(u_2^{(0)}, u_3^{(0)}\right)$. For small t, (10) agrees with (8), yielding linear growth for u_1, as expected; but for larger t, (10) rounds off and $u_1(x, t)$ never exceeds the largest value of $V_1^{(0)} - w$ along that streamline of (u_2, u_3) which passes through $\underline{x}$.

In case w is totally independent of $\underline{x}$, the coefficients multiplying t in (8) vanish identically, so that the linear growth fails to appear. Such a flow, however, is just a Galilean translate of the null flow, that is, no flow at all. All nontrivial, nondissipative, stationary parallel flows exhibit the spontaneous instability.

H. L. Willke, Jr.

C. DISPERSION RELATION FOR ULTRASOUND IN GYROTROPIC QUANTUM PLASMA

We report here a novel contribution to the slowly developing field theory of coupled electron-phonon systems.

The importance of the Feynman diagrammatic technique as a calculational tool in statistical physics is now probably well-known, and it requires only a slight stretch of the imagination to see that much more of the mathematical apparatus of quantum field theory (as it is now known), than just the Feynman diagram, can be used with advantage, mutatis mutandis, outside the context of elementary particle physics. We have in mind in this report, the method of single-variable dispersion relations.

We have used this technique in the study of the causal aspects of the problem of ultrasonic absorption in simple metals in a static magnetic field. Specifically, we have

derived the dispersion relation for the scattering of phonons from electrons in a magnetic field.

An interacting electron-phonon system is momentarily created when a beam of phonons, specified by a current vector $\vec{J}(\vec{r}, t)$ enters a solid, and eventually gets out of it, at a modified intensity. The external source of phonons is, typically, an ultrasonic wave generated by transducers in contact with the crystal. Then, under the assumption that this source establishes a single-frequency phonon spectrum in the solid, the attenuation of sound should be affected not only by single phonon processes, but also by multi-phonon processes.

In this report we discuss only forward dispersion relations. It has been found[1] that nonforward dispersion relations cannot be derived for this system.

We construct conjugate electron fields $\psi(\bar{x})$, $\bar{\psi}(\bar{x})$:

$$\psi(\bar{x}) = \sum_{p_z, s} a_s(p_z) \, \psi_{p_z, s}(\bar{x})$$

$$\psi(\bar{x}) = \sum_{p_z, s} a_s^\dagger(p_z) \, \psi_{p_z, s}^*(\bar{x})$$

where $\psi_{p_z, s}(\bar{x}) \equiv \phi(p_z, \bar{x}) \chi_s$ obeys the familiar equation

$$\left(\frac{\hbar^2}{2m} \left(\bar{p}_z + \frac{|e|}{c} \bar{A} \right)^2 + mc^2 + \sigma_z \frac{g}{4} \hbar \omega_H - E \right) \phi(p_z, \bar{x}) \chi_s = 0,$$

in which $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, $\chi_s = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ for spin up, and $\chi_s = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ for spin down. Here, a and $a^\dagger$ are annihilation and creation operators with respect to a vacuum, which is a full Fermi sea, they satisfy the usual anticommutation rules for fermions. We are using the free-electron model for electrons in metals.

The energy levels are given by

$$E_{n,a}(p_z) = \frac{p_z^2}{2m} + \left(n + \frac{1}{2} + \sigma_a \frac{g}{4} \right) \hbar \omega_H + mc^2$$

where $\sigma_a = 1$ for spin up, $\sigma_a = -1$ for spin down, and $n = 0$ or an integer. Assuming a coulombic electron-ion interaction, an electron-phonon interaction Hamiltonian, H_{int}, is readily derived in the form

$$H_{int} = \Omega_e \sum_{p_z' p_z, \bar{q}} \int d\bar{r} \, \bar{\psi}_{p_z'}(\bar{r}) \, \psi_{p_z}(\bar{r}) \frac{i}{\sqrt{V}} \left(-Q_\mu(\bar{q}) \, e^{-i\bar{q} \cdot \bar{r}} + Q_\mu^\dagger(\bar{q}) \, e^{i\bar{q} \cdot \bar{r}} \right)$$

$$\times \frac{m \hat{\epsilon}^\mu \cdot \hat{z} \, e^{-i\bar{K}_n \cdot \bar{r}}}{p_z' - p_z} \delta_{\bar{q}, \bar{K}_n - p_z' + p_z}$$

where $\Omega_e = \dfrac{4\pi Z e^2 N}{mV}$, Z is the ion valency, $\bar{K}_n$ is an arbitrary reciprocal lattice vector, and $Q_\mu(\bar{q})$ is the $\bar{q}^{th}$ Fourier component of the displacement of a lattice site and corresponds to a polarization μ ($\mu = 1, 2, 3$).

From H_{int}, we may define a free (vector) phonon field,

$$\bar{\phi}(\bar{r}) = i \sum_{\bar{q}, \mu} \frac{\omega(\bar{q}) \hat{\epsilon}^\mu}{\sqrt{V}} \left(-Q_\mu(\bar{q}) \, e^{-i\bar{q} \cdot \bar{r}} + Q_\mu^\dagger(\bar{q}) \, e^{i\bar{q} \cdot \bar{r}} \right)$$

and a "bare" coupling function

$$g_n^{\bar{q}, p_z' p_z}(\bar{r}) = \frac{\Omega_e}{\omega(\bar{q})} \frac{m \, e^{-i\bar{K}_n \cdot \bar{r}}}{(p_z' - p_z)} \delta_{\bar{q}, \bar{K}_n - p_z' + p_z}.$$

The latter we may, for simplicity, denote by $g(\bar{r})$. $\bar{\phi}(r)$ may be put into a second quantized form

$$\bar{\phi}(\bar{r}) = \sum_{\bar{q}, \mu} \hat{\epsilon}^\mu \sqrt{\frac{\omega(\bar{q})}{2}} \left(b_\mu(\bar{q}) \, e^{-i\bar{q} \cdot \bar{r}} + b_\mu^\dagger(\bar{q}) \, e^{i\bar{q} \cdot \bar{r}} \right),$$

in which the annihilation and creation operators $b_\mu(q)$, $b_\mu^\dagger(q)$ obey the usual commutation rules for bosons.

In terms of the new (unrenormalized) quantities, we have

$$H_{int} = \sum_{\bar{q}, p_z' p_z} \int d\bar{r} \left(\bar{\psi}_{p_z'}(\bar{r}) \, \hat{z} \, \psi_{p_z}(\bar{r}) \right) \cdot \bar{\phi}_{\bar{q}}(\bar{r}) \, g(\bar{r}) \equiv \int d\bar{r} \, H(\bar{r}).$$

Since there is no derivative coupling involved, this may be identified with the negative of an interaction Lagrangian, $L(\bar{r})$, from which we deduce our S-matrix,

$$S = T \left(e^{i \int_{-\infty}^{\infty} L(z) \, dz} \right),$$

where $z \equiv (\bar{z}, t)$ and all time translations are obtained by the unitary transformation

$$O(t) = e^{i(H_o - \mu N) t} \, O \, e^{-i(H_o - \mu N) t}$$

where H_0 is the unperturbed Hamiltonian, μ is the chemical potential, and N is the number of particles in the system.

In keeping with the spirit of this method, we now specify a number of fundamental postulates, among which are the existence of a stable vacuum (full Fermi sea), stable single particle states, a group of translations which transforms according to a certain unitary representation that also describes the transformations of the S-matrix and of its functional derivatives, a completeness relation for the combined electron-phonon states, and a causality condition expressed as

$$\frac{\delta}{\delta u(y)} \left(\frac{\delta S}{\delta u(x)} S^{\dagger} \right) = 0 \qquad \begin{array}{l} \text{for } t \equiv y_0 - x_0 < 0 \\[6pt] \text{or } (\bar{y} - \bar{x})^2 - c^2 t^2 > 0. \end{array}$$

This condition of causality can be shown[1] to be intimately connected with the physical concepts of renormalization, according to which, in the viewpoint that we adopt here, the free fields $\psi(\bar{x})$, $\phi(\bar{x})$, defined above, and the electron-phonon vertex $\Gamma(\bar{x})$ are transformed:

$$\psi(\bar{x}) \rightarrow \frac{\psi(\bar{x})}{\sqrt{Z_2(\bar{x})}} \; ; \qquad \phi(\bar{x}) \rightarrow \frac{\phi(\bar{x})}{\sqrt{Z_3(\bar{x})}} \; ,$$

and $\Gamma(\bar{x}) \rightarrow Z_1(\bar{x}) \Gamma(\bar{x})$, when and only when the electron-phonon interaction is "switched on." The function $g(\bar{r})$ is transformed as $g(\bar{r}) \rightarrow Z_2 Z_1^{-1} Z_3^{1/2} g(\bar{r})$, and all conceivable infinities are eliminated.

We consider the process shown schematically in the following diagram.
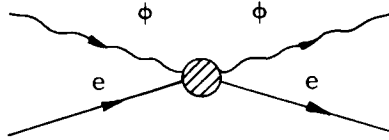
The scattering amplitude which describes this process may be defined as

$$\left\langle p_z' s'; \bar{q}' \mu' \mid S \mid p_z s; \bar{q}\mu \right\rangle = \delta(p_z' - p_z) \, \delta_{ss'} \delta(\bar{q}' - q) \, \delta_{\mu\mu'}$$

$$+ \frac{1}{2} \sqrt{\omega(\bar{q})\omega(\bar{q}')} \; \delta(p_z' + \bar{q}' - p_z - \bar{q}) \; \delta(E(p_z') + \omega(\bar{q}') - E(p_z) - \omega(\bar{q}))$$

$$\times f_{s', \mu'; s, \mu}(p_z' \bar{q}'; p_z \bar{q}); \qquad s \equiv \{\text{all discrete parameters}\}$$

By means of a contraction scheme involving the commutators of the S-matrix with the various creation and annihilation operators, we may reduce the matrix element $\left\langle \text{final} \mid S \mid \text{initial} \right\rangle$ to one between electron states only. We may then proceed to relate the scattering amplitude to two auxiliary functions F^a, F^r, which are of the nature

of "advanced" and "retarded" functions, respectively.

This is a convenient way of glossing over the complicated algebra which brings us to the following remarkably simple results:

(1) The amplitude for phonon absorption exceeds the amplitude for phonon emission in the ratio 3 to 1 for forward scattering, so there is an over-all ultrasonic attenuation.

(2) Two forward-dispersion relations may be written for the electron-phonon system. The first one, which is similar to the Kramers-Kronig relation for light waves, is

$$\text{Re } f(\omega) = g^2 \hbar/mc + \frac{P}{2\pi^2 c} \int_0^\infty \frac{\sigma(\omega') \, \omega'^2}{\omega'^2 - \omega^2} \, d\omega'$$

and is independent of the magnetic field. Here, $\sigma(\omega)$ is the cross-section, ω is the phonon energy variable, and g^2 is the electron-phonon coupling constant. The second dispersion relation is magnetic field-dependent and is

$$\text{Re } f(\omega) = g^2 \hbar/mc + \frac{P}{2\pi^2 c} \int_{\omega_1}^\infty \frac{\sigma(\omega') \, \omega'^2 \, d\omega'}{\omega'^2 - \omega^2}$$

where $\omega_1 = \dfrac{2mS^2}{\cos^2 \theta}$, S is the unrenormalized speed of sound, and θ is the inclination of sound wave vector to magnetic field in the z-direction.

Of the five possible ultrasonic absorption phenomena in solids, geometric, cyclotron, and open-orbit resonances, de Haas-van Alphen type, and giant quantum oscillations, we are of the opinion that the second dispersion relation describes giant quantum oscillations, while the first describes all of the other ordinary absorption phenomena.

A number of deductions could be made from this relation:

(a) There is a nonzero frequency threshold for the onset of absorption of the giant quantum type.

(b) This threshold is a minimum for propagation parallel to the field and recedes to infinity as propagation approaches the transverse direction. Therefore we cannot expect to find such oscillations for transverse propagation at any frequency.

(c) Knowing this threshold, we may determine the value of the parameter, m, which is of the form of an effective mass. Numerical estimates with $\omega_1 \sim 10$ Mc/sec, give $m \sim 0.01$ times the free electron mass, which appears reasonable.

(d) Since ω_1 is nonzero, we may introduce the concept of an effective phonon mass in a magnetic field. This mass, $\mu(\theta)$, is dependent on the propagation direction.

<div align="right">A. A. Maduemezia, K. U. Ingard</div>

<div align="center">References</div>

1. A. A. Maduemezia (unpublished thesis research).

D. DIFFUSION WAVES

Consider a simplified model of a weakly ionized gas, in which the electrons are assumed to diffuse instantaneously ($T_e \rightarrow \infty$) and form a uniform negative background. Then, if the ion mobility is neglected, and the first-order perturbation in the ionization rate Z is proportional to the perturbation in the electric field E, the ion continuity equation is

$$n_t = \gamma N e \tag{1}$$

where n is the perturbation in ion density, e is the perturbation in electric field, $\gamma = (dZ/dE)_o$, and N is the unperturbed density. Also, we use the Poisson equation

$$e_x = qn \tag{2}$$

where q is the electronic charge.

If we set

$$t' = t$$

$$x' = \gamma q N x$$

$$n' = n/N \tag{3}$$

$$e' = -\gamma e$$

and drop the primes, Eqs. 1 and 2 become

$$n_t = -e; \qquad e_x = n. \tag{4}$$

(The negative sign was inserted to give wave motion in the positive x-direction.) Equation 4 may be written as a single equation for n (or e),

$$n_{tx} + n = 0. \tag{5}$$

With a solution of the form $\exp(ikx-i\omega t)$, Eq. 5 gives a dispersion relation

$$\omega = -1/k, \qquad V_{ph} = \omega/k = -1/k^2, \qquad V_{gr} = d\omega/dk = 1/k^2. \tag{6}$$

The group and phase velocities are opposite in sign, as can be seen by examining a wave-packet solution of Eq. 5. The general solution of Eq. 5 is

$$n(x,t) = \int_{-\infty}^{\infty} N(k) \exp(ikx+it/k) \, dk. \tag{7}$$

We choose $N(k) = (1/\sqrt{\pi}) \exp\left(-(k-k_o)^2/4\right)$ where $k_o \gg 1$, and evaluate the integral in Eq. 7 by expanding the exponential about $k = k_o$ using the saddle-point method. Then

$$n(x,t) = \exp\left(-\left(x-t/k_o^2\right)^2\right) \exp(ik_o x + it/k_o). \tag{8}$$

This solution is presented graphically in Fig. IX-22 for $k_o = 4$. We see that the individual waves propagate in the -x direction, while the wave packet propagates in the +x direction.
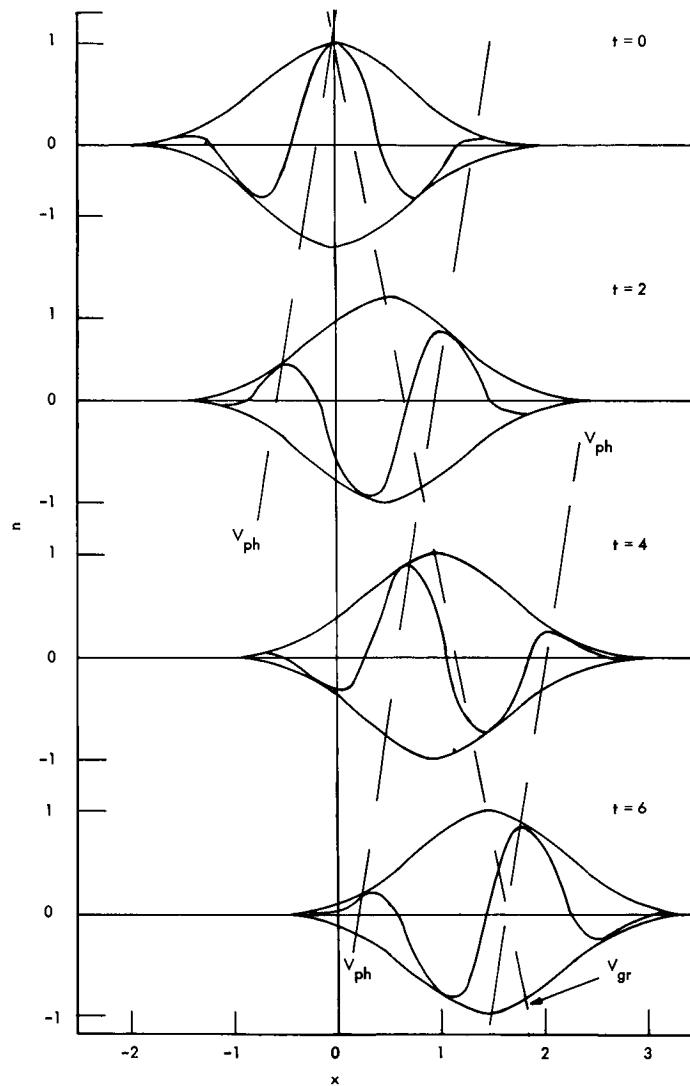
Fig. IX-22. Propagation of wave packet.

Some other properties of diffusion waves may be found by studying Eq. 5. This equation may be derived from the Lagrangian

$$L = n_t n_x - n^2. \tag{9}$$

Calculation of the stress-energy tensor components[1] gives

canonical momentum $= n_x$

"stress" density $= n^2$

"energy" flow $= n_t^2$ (10)

wave momentum density $= n_x^2$

"energy" density (Hamiltonian) $= n^2$.

The divergence of the stress tensor gives an "energy" conservation equation which enables us to find an "energy" flow velocity.

$$(n^2)_t + \left(n_t^2\right)_x = 0 \tag{11}$$

For a wave moving with the energy flow velocity U,

$$d/dt = U \ d/dx;$$

thus

$$\left(n_t^2 + U n^2\right)_x = 0 \tag{12}$$

and

$$U = -n_t^2/n^2, \tag{13}$$

which is the group velocity $V_{gr} = \omega^2$.

The waves produced by an impulse at $t = 0$ show a similarity to certain electromagnetic waves. To illustrate this, we use Eq. 7. Since

$$n(x, 0) = \int_{-\infty}^{\infty} N(k) \ \exp(ikx) \ dk, \tag{14}$$

if we choose $N(k) = 1/2\pi$, then $n(x, 0) = \delta(x)$. If we define

$$kx + t/k = \sqrt{tx} \left(k\sqrt{x/t} + \frac{\sqrt{t/x}}{k}\right) = 2\sqrt{tx} \cos u \tag{15}$$

where $\exp(iu) = k\sqrt{x/t}$ and $dk = i\sqrt{t/x} \exp(iu) \ du$, Eq. 7 becomes

$$n(x, t) = (i/2\pi) \sqrt{t/x} \int_0^{2\pi} \exp(2i\sqrt{tx} \cos u) \exp(iu) \ du = -\sqrt{t/x} \ J_1(2\sqrt{xt}), \tag{16}$$

which is shown graphically in Fig. IX-23. The fact that wave motion is not observed in
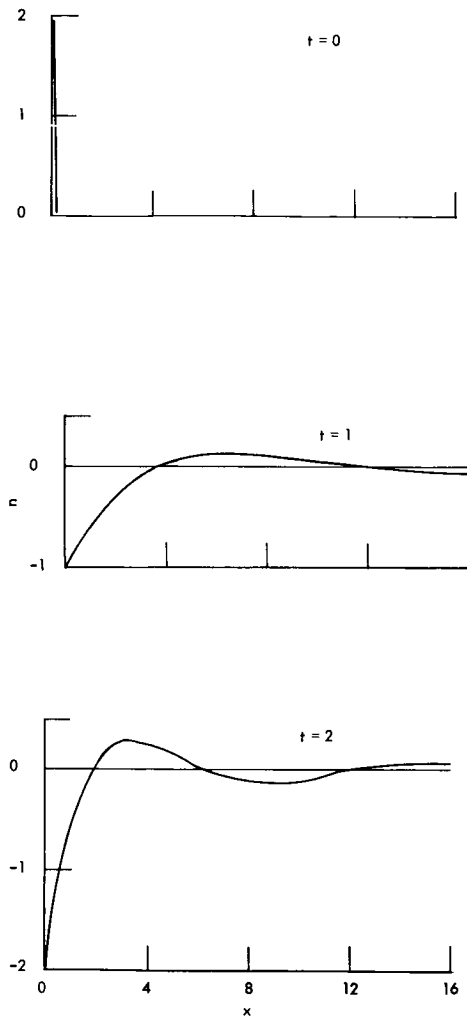
Fig. IX-23. Propagation of impulse.

the -x direction may be expected because Eq. 5 is not symmetric in x. Figure IX-23 shows that, although the disturbance propagates away from the origin, the individual waves move toward the origin. Thus the first zero of n is at x = 4 for t = 1 and is at x = 2 for t = 2. It is also evident that the long-wavelength components of n propagate faster than the short-wavelength components, as one would expect from the dispersion relation.

The results of Eqs. 14-16 are also obtained for the case of "forerunner" light waves,[2] which travel with velocity c in a dispersive medium. These light waves are called forerunner waves because they are received before the main signal, which travels at the group velocity, which is less than c. The forerunners are illustrated in Fig. IX-23 in a coordinate system moving with velocity c.

The characteristic curves for Eq. 5 are x = const and t = const. This suggests that, by rotating the coordinates 45°, Eq. 5 can be brought into the form of a wave equation. If we set $\psi = x + t$, and $\phi = x - t$, Eq. 5 becomes

$$n_{\psi\psi} - n_{\phi\phi} - n = 0, \tag{17}$$

which is the Klein-Gordon or Telegrapher's equation. The Green's function for Eq. 5 is found to be

$$G(x, t; x_0, t_0) = J_0(2\sqrt{(x-x_0)(t-t_0)}) \tag{18}$$

by transforming the Green's function for Eq. 17. In formulating an initial-value problem for Eq. 5, we cannot specify both n(x, 0) and $n_t(x, 0)$, since t = 0 is a characteristic curve.

There are several known examples of materials with group and phase velocities in opposite directions. Lamb[3] gives the case of a wire that is subjected to a longitudinal thrust (rather than tension) and to a linear restoring force. The equation of motion for transverse vibrations (in suitable units) is

$$y_{tt} + y_{xx} + y = 0.$$ (19)

These vibrations have a dispersion relation

$$\omega^2 = 1 - k^2.$$ (20)

Since

$$d(\omega^2)/d(k^2) = V_{gr}V_{ph} = -1,$$ (21)

it is seen that V_{gr} and V_{ph} are in opposite directions. The dispersion relation, Eq. 20, is, however, considerably different from the dispersion relation, Eq. 6. Lamb also refers to a model discussed by Rayleigh[4] of a wire under tension, which has a linear restoring force, rotatory inertia, but negligible stiffness. Lamb states that this model may apply to "a cylindrical wire with a series of close equidistant peripheral cuts extending nearly to the axis." If we neglect the tension, the equation of motion is

$$y_{tt} + y - y_{xxtt} = 0,$$ (22)

and the dispersion relation is

$$\omega^2 = 1/(1+k^2),$$ (23)

which is the same as Eq. 6 when k is large.

It is clear that the dispersion relation $\omega = -1/k$ cannot be valid for very small or very large k. For k very large (short wavelength), the dispersion relation predicts very small wave speeds. This implies that steep gradients in n tend to remain stationary. This result is due to the neglect of pressure-gradient forces. When pressure is included, the phase and group velocities both approach the speed of sound, when k is large. For very small k (long wavelength), the dispersion relation predicts very large wave speeds. As small k implies large ω, these waves involve very rapid changes with time. With these very rapid changes, it is no longer correct to neglect inertial forces and assume that the electrons diffuse instantaneously. With these corrections made, both the phase and group velocities remain finite for small k.

It is these additional terms that make diffusion waves of interest in the study of gas discharges. With the coupling between sound waves and diffusion waves included, the dispersion relation takes the form[5]

$$\omega = -\frac{1}{2k} \pm \sqrt{\left(\frac{1}{2k}\right)^2 + k^2}$$ (24)

where one mode has zero group velocity for $k \approx 1$. These results are more complicated for finite electron temperatures.

S. D. Weiner, U. Ingard

References

1. P. M. Morse and H. Feshbach, Methods of Theoretical Physics, Vol. 1 (McGraw-Hill Publishing Company, New York, 1953), p. 305.

2. L. Brillouin, Wave Propagation and Group Velocity (Academic Press, New York, 1960), p. 41.

3. H. Lamb, On group velocity, Proc. London Math. Soc. 1, Series 2, 473 (1904).

4. Lord Rayleigh, Scientific Papers, Vol. IV (Cambridge University Press, London, 1892-1901), p. 369.

5. S. D. Weiner and U. Ingard, Quarterly Progress Report No. 76, Research Laboratory of Electronics, M. I. T., January 15, 1965, pp. 72-75.

PLASMA DYNAMICS

X. PLASMA PHYSICS[*]

Prof. S. C. Brown
Prof. W. P. Allis
Prof. G. Bekefi
Prof. K. U. Ingard
Prof. D. R. Whitehouse
Dr. J. C. Ingraham
Dr. G. Lampis
M. L. Andrews
F. X. Crist

J. K. Domen
E. W. Fitzgerald, Jr.
G. A. Garosi
K. W. Gentle
E. V. George
W. H. Glenn, Jr.
E. B. Hooper, Jr.
P. W. Jameson
R. L. Kronquist
D. T. Llewellyn-Jones

E. M. Mattison
J. J. McCarthy
W. J. Mulligan
J. J. Nolan, Jr.
L. D. Pleasance
G. L. Rogoff
D. W. Swain
F. Y-F. Tse
B. L. Wright

A. ON THE AMBIPOLAR TRANSITION

A report[1] by Cohen and Kruskal (referred to here as C+K) considers the ambipolar transition by dividing the problem into a number of "regions" in which the equations may be simplified, and this is very helpful in understanding the problem. The report is, however, difficult to follow for a number of reasons, and it therefore appears useful to write the present report, containing little that is factually different from theirs, but in which some more complete equations are presented and the structure of the problem is more carefully exhibited. The method of C+K makes mathematical rigor possible, which is excellent, but the report is not convincing that the rigor is always there. Hidden errors may lurk where the interrelations of equations are not clearly shown. C+K consider various limiting forms assumed by the complete ambipolar diffusion equations when certain parameters go to zero or infinity. In each case the limiting process results in dropping one term in each of several three-term equations. In any physical problem where parameters and variables may have values ranging over orders of magnitude it is generally true that, to some approximation, the smallest term out of three may be neglected. In this problem the three terms A, B, C, will be of the same sign, so that the full equation can be written

$$A + B = C \tag{AB}$$

with two simplified forms

$$A = C \tag{A}$$

or

$$B = C \tag{B}$$

These simplified forms will be called "limits" in the sense that some physical process has become inoperative when a term can be dropped. This is different from the mathematical meaning of "limit" used by C+K, although the result is generally the same. The

---

full equation (AB) will be called a "transition" and must be used in the neighborhood of A = B. In a "transition" two physical processes are competing.

As there will be 3 three-term equations, there are 8 possible limits, but as we shall assume that the positive ions are both heavier and colder than the electrons, only 4 of the limits actually occur. They will be named, somewhat arbitrarily, the Ambipolar (a), Boltzmann (b), Cosine (c), and Diffusion (d) limits. These will be discussed in reverse order. They are separated by the Space-Charge (ab), Electron-Flow (bc), and Ion-Flow (cd) transitions. We also consider the sheath limit (s) in which ionization is neglected.

It must not be assumed that the same limit holds throughout a given plasma. In general, the transition condition A = B will cut across the diffusing, and therefore nonuniform, plasma so that several limiting and transition forms must be used. It is our purpose to sketch the way in which these limiting and transition forms fit together in a given plasma. This leads to our final diagram (Fig. X-3) which summarizes as much information as can be collected in a single figure.

In order to facilitate comparison with the work of C+K their equation number (D. _. _) or the page reference (Eq. _ p. _) will be given next to ours whenever we have found the appropriate reference.

(e) Basic Equations

The physical parameters of this problem are the ion and electron temperatures $T_\pm$ and transport coefficients $\mu_\pm$, $D_\pm = \mu_\pm T_\pm$, the ionization frequency ν_i, and the diffusion length $\Lambda = L/\pi$. Parallel plane geometry will be considered. These physical parameters are combined in the following dimensionless parameters:

$$\tau = T_+/T_- \leqslant 1$$

$$\mu = \mu_+/\mu_- < 1$$

$$\delta = D_+/D_- = \tau\mu \ll 1$$

(e 1) (p. 15)

$$\nu = \nu_i \Lambda^2/T_- \mu_+$$

and use will also be made of the electric field $\mathscr{E}_i$ defined as

$$\mathscr{E}_i^2 = \nu_i T_-/\mu_+ .$$

(e 2)

(Note that our τ is $1/\tau$ in C+K.) The first three parameters are less than 1, the fourth is a characteristic value obtained from the solution of the equations. Its free and ambipolar limits are $1/\mu$ and $(1+\tau)/(1+\mu)$; C+K claim that this last limit is approached from below. We claim this has not been proved.

We then introduce the dimensionless variables

$$n_{\pm} = N_{\pm} e \mu_+ / \epsilon_o \nu_i$$

$$s = n_+ - n_-$$

$$E = \mathscr{E} / \mathscr{E}_i \qquad \text{(e 3) (p. 15)}$$

$$J = \Gamma_e / \epsilon_o \mathscr{E}_i \nu_i$$

$$x = X \mathscr{E}_i / T_-$$

and use the same variables throughout. They differ slightly from the variables used by Allis and Rose[2] (referred to here as A+R). C+K rescale variables on pp. 33, 37, 38, 45, 59, 60, 71 and this makes their equations difficult to compare. Our single set of variables resembles their barred variables but differs from them by factors of $\pi/2$ because of the use of the diffusion length Λ in the dimensionless parameters (e 1). Ours are chosen so that τ and μ tag those terms in the equations which are going to be neglected and that ν not appear explicitly in the equations. They are all of order 1 at the space-charge transition.

In terms of these variables, the ion flow, electron flow, Poisson and ion generation equations are

$$J = -\tau \nabla n_+ + E n_+$$

$$\mu J = -\nabla n_- - E n_-$$

$$\nabla \cdot E = n_+ - n_- = s \qquad \text{(e 4) (B. 1)}$$

$$\nabla \cdot J = n_-$$

with the boundary conditions

$$n_- = n_o \qquad n_+ = n_{+o} \qquad \text{at } J = E = x = 0$$

$$J = J_d, \quad E = E_d, \quad x = x_d \qquad \text{at } n_- = n_+ = 0. \qquad \text{(e 5)}$$

The reason for the subscript d at the wall will appear later. Here, n_o is quite arbitrary but n_{+o}/n_o is determined a posteriori by requiring that n_+ and n_- go to zero at the same place. J_d, E_d, and x_d have whatever values result from the solution. By requiring that x_d correspond to $X = L/2$ at the wall, it follows from (e 1) and (e 3) that in these variables

$$\sqrt{\nu} = 2 x_d / \pi. \qquad \text{(e 6)}$$

Thus the characteristic value is computed directly without the need of any fitting.

The set of equations (e 4) is assumed to be valid all the way to the wall, which implies that the collision mean-free path is much smaller than the thickness of any boundary

layer. This is rarely true, but the assumption provides definiteness to the mathematical problem.

Because the equations are invariant to translations in x, it is convenient to eliminate it as independent variable in favor of the current, J (p. 38). This is done by use of the ion-generation equation and leads to the set of 3 three-term equations mentioned above. Primes are used for derivatives with respect to J.

$$-\tau n_- n_+' = J - En_+$$

$$-n_- n_-' = \mu J + En_- \qquad\qquad (e\ 7)$$

$$n_- E' = n_+ - n_-$$

The space coordinate is then obtained by quadrature

$$x = \int_0^{} \frac{dJ}{n_-} = \int^{n_0} \frac{dn_-}{\mu J + En_-}. \qquad\qquad (e\ 8)$$

Certain exact relations can be derived from the set of equations (e 7), and in order to be perfectly clear as to which equations are exact, they are all summarized here. Limit solutions can then be compared with them to see explicitly what has been neglected. Taking the derivative of the ion flow equation and combining with Poisson's equation gives

$$\left(\frac{n_+}{n_-} - 1\right) n_+ = 1 - En_+' + \tau(n_- n_+'). \qquad\qquad (e\ 9)$$

Combining the equations in different ways yields the two relations that are useful near the ambipolar limit

$$(1-\delta)J - (1+\tau)En_- = Es - \tau n_- s' \qquad\qquad (e\ 10)\ (D.\ 3.\ 5)$$

$$(1+\mu)J + (1+\tau)n_- n_-' = Es - \tau n_- s' = \sigma. \qquad\qquad (e\ 11)\ (D.\ 3.\ 8)$$

Eliminating J from the flow equations, combining with Poisson's equation, and integrating gives

$$\frac{1}{2}\mu E^2 + (1+\mu) \int_0^{} E\ dJ = n_0 - n_- - \delta(n_{+0} - n_+). \qquad\qquad (e\ 12)$$

A somewhat different combination of the equations yields, upon integrating,

$$\left(J + \frac{\mu E}{1+\mu}\right)^2 = \frac{1+\tau}{1+\mu}(n_0 n_{+0} - n_- n_+) + \frac{2\mu}{1+\mu}[n_0 - n_- - (n_{+0} - n_+)] + \frac{1-\tau}{1+\mu}\int^{n_0}(n_+\ dn_- - n_-\ dn_+). \qquad\qquad (e\ 13)$$

For an isothermal plasma $\tau = 1$, so that the integral in (e 13) drops out and the remaining terms are a first integral of the equations.

Finally, n_- can be divided out of (e 11) so that it integrates with respect to $dJ = n_- dx$.

$$(1+\mu) \int_0^{} J \, dx = E^2/2 + n_o - n_- + \tau(n_{+o}-n_+) \tag{e 14}$$

$$(1+\mu) \int_0^{Xd} J \, dx = E_d^2/2 + n_o + \tau n_{+o}. \tag{e 15}$$

(d) <u>Diffusion Limit</u> $\mu_- \to \mu_+ \to 0$ $n_+ \ll \delta$

In the limit of very low densities the quadratic terms En_+ and En_- can be neglected so that the limiting equations are
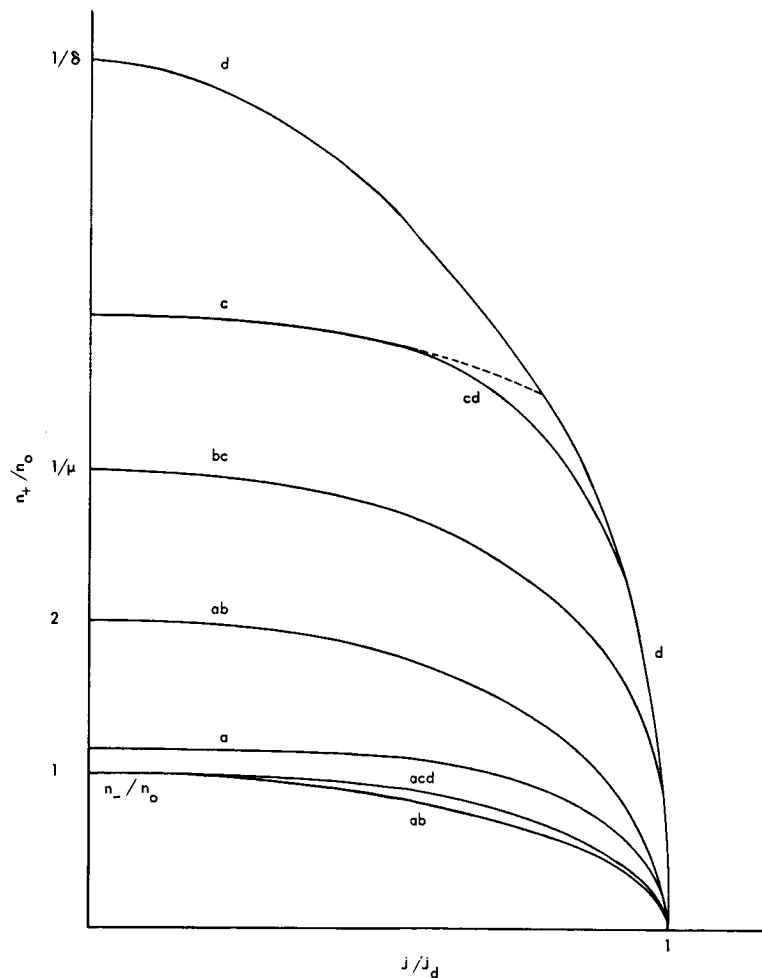
Fig. X-1.   Plasma profiles.

$$-\tau n_- n_+' = J$$

$$-n_- n_-' = \mu J \qquad \text{(d1)}$$

$$n_- E' = n_+ - n_-,$$

and the solution is

$$\delta n_+ = n_-$$

$$\delta E = (1-\delta)J \qquad \text{(d2) (D. 4. 3)}$$

$$\mu J^2 = n_o^2 - n_-^2$$

$$n_- = n_o \cos \sqrt{\mu}\, x \qquad \text{(d3)}$$

$$\nu = \nu_d = 1/\mu.$$

It will be convenient to plot n_-/n_o and n_+/n_o against J/J_d (see Fig. X-1). Both plots are ellipses in this limit and remain identical for all sufficiently small values of n_o. As C+K show quite generally that

$$\delta n_+ \leqslant n_- \leqslant n_+, \qquad \text{(d4) (B. 6)}$$

all future curves of n_+/n_o will fall between these two limit curves.

Putting the value of E from (d2) into the electron flow equation (e 7) yields the first-order departure

$$\sqrt{\mu\nu} = 1 - \frac{2}{3\pi}\, \frac{1-\delta}{\delta}\, \frac{n_o}{\mu}. \qquad \text{(d5) (p. 75)}$$

(cd) <u>Ion Flow Transition</u> $\mu_- \to n_-/n_+ \to 0$ $n_{+o} \approx \delta$

As $En_-/\mu En_+ \approx \tau$, and we assume $\tau < 1$, the ion mobility term will, with increasing density, become equal to and cross the ion diffusion term before the electron mobility term need be considered. In an active plasma the electric field assists the ions to flow out before it effectively retains the electrons. This is the Ion Flow Transition. It occurs before the Electron Flow Transition in an active plasma, although they occur together in an isothermal plasma. We shall consider them separately. The appropriate equations for the Ion Flow Transition are

$$-\tau n_- n_+' = J - En_+$$

$$-n_- n_-' = \mu J \qquad \text{(cd1)}$$

$$n_- E' = n_+.$$

From (e 12) we see that

$$\frac{1 + \tau}{2} \mu E^2 = n_o - n_- - \delta(n_{+o} - n_+),$$ (cd 2) (D. 2. 4)

and from the electron flow equation, as before,

$$\mu J^2 = n_o^2 - n_-^2$$
$$\mu \nu = 1.$$ (cd 3)

(c) Cosine Limit $D_+ \rightarrow n_-/n_+ \rightarrow 0$ $\delta < n_{+o} < \mu$

After the ion flow transition, the appropriate equations are

$$J = En_+$$
$$-n_- n_-' = \mu J$$ (c 1)
$$n_- E' = n_+.$$

The electron profile is the same as before

$$\mu J^2 = n_o^2 - n_-^2$$
$$\mu \nu = 1,$$ (cd 3)

but now we have

$$\frac{1}{2} \mu E^2 = n_o - n_-$$ (D. 2. 8)*

$$n_+^2 = (n_o + n_-)/2$$ (c 2)

$$n_+ = \sqrt{n_o} \cos \sqrt{\mu} \, x/2.$$ (D. 2. 7)

[*In going from (D. 2. 4) to (D. 2. 8) C+K neglect δn_{+o} but keep δn_+. This seems inconsistent.]

This limit has been called the Cosine Limit because n_+ is a cosine function; but it is of twice the base, so that it does not vanish at the wall. Clearly this limit, as all of the regions that follow, are not valid beyond the point where their plasma profiles intersect the curve (d) of Fig. X-1. The solutions (c 2) must be broken off at a coordinate $x_c < x_d$ at which the conditions of validity of (c 1) break down and a single new boundary condition of the form

$$n_-/n_+ = c$$ (c 3)

must replace (e 5).

The impossibility of satisfying the double boundary condition $n_+ = n_- = 0$ arises because in dropping the ion diffusion term the order of the equations has been reduced by one, and hence there is one less disposable constant of integration. This becomes obvious when we drop the diffusion term in (e 9). This then becomes

$$\frac{n_+}{n_-} = 1 + \frac{1 - En_+'}{n_+} \geq 1 + \frac{1}{n_+} \qquad (c\ 4)$$

and

$$\frac{n_{+o}}{n_o} = 1 + \frac{1}{n_{+o}}. \qquad (c\ 5)$$

Thus n_o and n_{+o} are related by (c 5) and only one of them may be chosen arbitrarily. Thus we no longer have the flexibility to require both n_+ and n_- to go to zero. In fact, neither one can go to zero before the conditions of the equation set (c 1) become invalid. (The appropriate constant in (c 3) will be discussed below in (s) Sheaths.) We also have a bound on n_+/n_- and it is now convenient to represent the solutions as plots of n_-/n_+
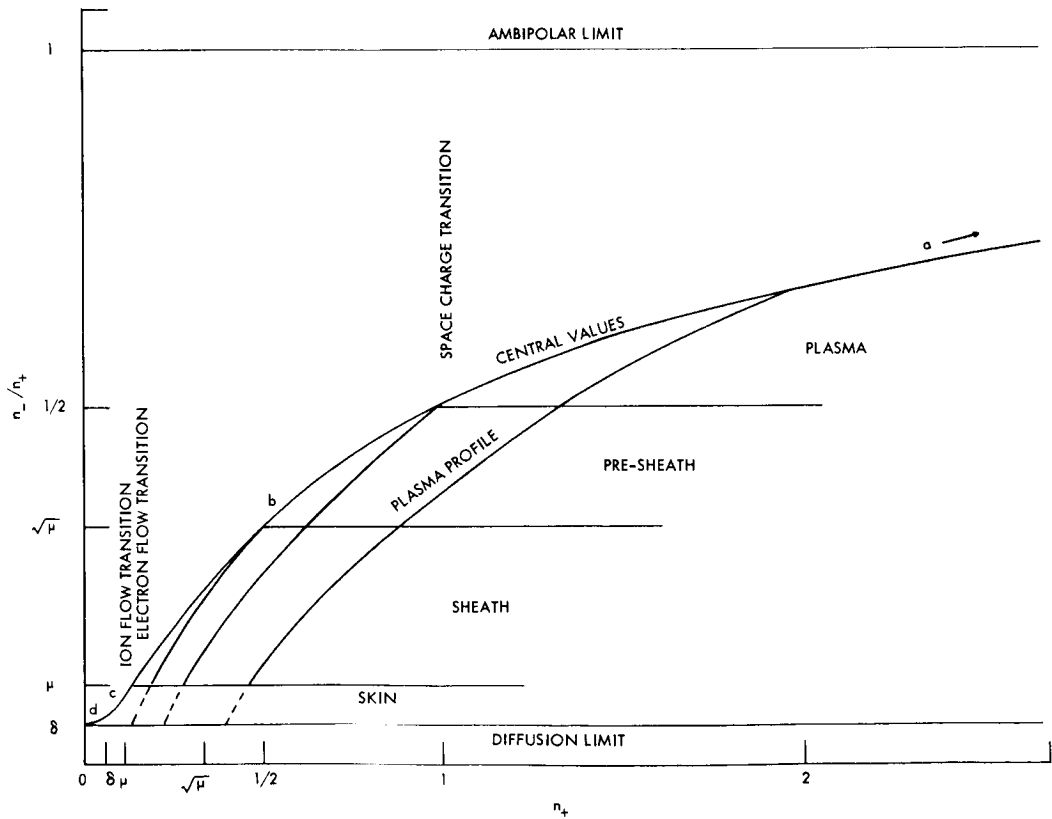

Fig. X-2. Plasma ratios.

against n_+ (Fig. X-2). For $n_+ > \delta$ all of the integral curves start on the hyperbola (c 4) at $n_+ = n_{+o}$ and remain below it. Eventually they curve sharply to the left, because of the ion diffusion term, and meet the axis of ordinates at $n_-/n_+ = \delta$. In practice, it may often be sufficient simply to cut the curves off at $n_-/n_+ = \delta$ without computing the ion diffusion part.

Note that (c 4) and (c 5) remain valid through (c), (b), and (a). Similarly, the electron profile (cd 3) remained valid through (d) and (c). The formula (d 5) produces a hardly noticeable change in ν through these regions. After (c) it is not the appropriate formula. We should now take E from (c 2) and put it into the electron-flow equation to get first-order departures for ν. This yields

$$\sqrt{\mu\nu} = 1 - \frac{2.13}{\pi} \frac{\sqrt{n_o}}{\mu}, \tag{c 6}$$

where the decimal comes from a complicated logarithmic term. This differs markedly from the previous formula (d 5) and is valid for considerably larger values of n_o.

(bc) __Electron-Flow Transition__ $\qquad D_+ \to n_-/n_+ \to 0 \qquad n_{+o} \approx \mu$

The next term to enter the equations is the electron mobility, which acts to oppose the electron flow. The equations then are

$$J = En_+$$

$$-n_- n_-' = J + En_- \tag{bc 1}$$

$$n_- E' = n_+.$$

A relatively simple equation can be obtained for the current squared $J^2 = F$ in terms of the field squared $E^2 = Y$. The equation is

$$2F'' + \sqrt{\frac{Y}{F}}\, F' + \mu = 0, \tag{bc 2} (D. 2. 11)$$

where the primes are derivatives with respect to Y, with

$$n_+ = \sqrt{F/Y}$$

$$n_- = F'.$$

A similar equation has been solved by C+K.

(b) __Boltzmann Limit__ $\qquad D_+ \to n_-/n_+ \to 0 \qquad \mu \to \infty \qquad \mu < n_{+o} < 1$

In this limit the ions are cold, flow is due to the field only, and the electrons have such large mobility that they assume a Boltzmann distribution in the field

$$n_- = n_0 e^{-V}. \tag{b1}$$

The equations are

$$J = En_+ $$

$$-n_-' = E \tag{b2}$$

$$n_- E' = n_+ $$

and they lead to the simple equation

$$J = nn'n''. \tag{b3} \quad \text{(D. 1. 9)}$$

This limit is called the "Similarity Solution" by C+K because it has an obvious scaling law:

$$
\begin{array}{ll}
n_- \rightarrow k^4 n_- & J \rightarrow k^3 J \\[2mm]
n_+ \rightarrow k^2 n_+ & E \rightarrow kE \\[2mm]
x \rightarrow s/k & \nu \rightarrow \nu/k^2 \\[2mm]
N_+ = \text{const} & N_- \nu_i = \text{const.}
\end{array}
\tag{b4} \quad \text{(Eq. 3 p. 43)}
$$

Here, the last line makes use of definitions (e 1) and (e 3). This implies that when N_0 varies in the range of validity of (b), which is two decades in Fig. 5 of C+K, ν_i varies inversely with N_0, and N_+ does not change at all! Fortunately there is a fallacy in this argument. The boundary condition (c 3) does not satisfy the scaling law. Hence x_c does not scale with x, and nothing can be said about the scaling of ν. The result on p. 43 of C+K is difficult to understand.

(ab) <u>Space Charge Transition</u> $\qquad D_+ \rightarrow 0 \qquad \mu_- \rightarrow \infty \qquad n_{+0} \approx 1$

In the regions considered thus far with $n_+ < 1$ it follows from (c 4) that $n_-/n_+ < 1/2$ and hence the electron space charge could, in the approximation used throughout this report, be neglected. It must now be included and the equations are

$$J = En_+ $$

$$-n_-' = E \tag{ab1}$$

$$n_+ = n_-(1+E') = n_-(1-n_-'') \tag{D. 1. 7}$$

which yield

$$J = n_- n_-' (n_-''-1). \tag{ab2} \quad \text{(D. 1. 8)}$$

The electron profiles plotted against x are bell-shaped with an inflection at $E^2 = s$. As E is large near the wall, the profile against J comes in to the axis nearly vertically, so that its departure from ellipticity is unnoticeable.

Note that the equations now contain no physical parameters. They are universal, and tabulated solutions with n_o as a parameter would be very useful; it is left to the user to join these solutions to the boundary solutions appropriate to his problem.

(a) <u>Ambipolar Limit</u> $n_+/n_- \to 1$ $n_o > 1$

In this limit the Poisson term is neglected. As the electron current and ion diffusion terms do not create extra difficulties, it is convenient to re-introduce them. The equations, therefore, are

$$-\tau n_- n_+' = J - En_+$$

$$-n_- n_+' = \mu J + En_- \qquad\qquad \text{(a 1) (D. 3. 2)}$$

$$n_+ = n_-$$

and the solution is

$$(1-\delta)J = (1+\tau)En_-$$

$$(1-\mu)J^2 = (1+\tau)\left(n_o^2 - n_-\right) \qquad\qquad \text{(a 2) (pp. 44 + 45)}$$

$$n = n_o \cos \sqrt{\frac{1+\mu}{1+\tau}}\, x \qquad\qquad \text{(D. 3. 3)}$$

$$v = v_a = \frac{1 + \tau}{1 + \mu}.$$

Computing the Poisson term, one finds

$$n_- E' = \frac{1-\delta}{1+\mu}\left(\frac{n_o}{n_-}\right)^2 = s. \qquad\qquad \text{(a 3)}$$

For equations (a 1) to be valid one must have

$$s < n_- \qquad\qquad \text{(a 4)}$$

or

$$n_-^3 > n_{ab}^2 = n_o^2 \qquad\qquad \text{(Eq. 3, p. 45)}$$

and the ambipolar limit is perhaps best expressed as

$$\text{limit } n_-^3/n_o^2 \to \infty. \qquad\qquad \text{(a 5)}$$

The ambipolar transition takes place first at the center and works outward, never reaching the wall.

From (a 3) we see that there is a space charge $s_0 = (1-\delta)/(1+\mu)$ at the center of the plasma and the space charge at the wall is much larger, although (a 3) is clearly excessive. It is interesting to know how large the effect of the distributed space charge is compared with that in the sheath.

Making use of (a 3) and neglecting terms in s^2/n_-^2, one finds that the right-hand side of (e 11) is

$$\sigma = -(1+\tau)\alpha s n',$$

where

$$\alpha = \frac{1 - 2\tau - 3\delta}{(1+\tau)(1+\mu)},$$

(a 6)

so that

$$(1+\mu)J^2 = (1+\tau)(n_0 + n_- + 2\alpha\bar{s})(n_0 - n_-)$$

(a 7)

and $\bar{s}$ is a number between s_0 and s.

Treating s as constant in (a 6) and integrating dJ/n_- from n_0 to n_a gives

$$\frac{x}{\sqrt{\nu_a}} = \frac{\pi}{2} - \sin^{-1}\frac{n_a + \alpha\bar{s}}{n_0 + \alpha\bar{s}} + \frac{\alpha s}{n_0}\ln\frac{2n_0}{n_a}$$

$$\rightarrow \frac{\pi}{2} - \frac{n_a}{n_0} + \frac{\alpha s}{n_0}\ln\frac{2n_0}{e n_a},$$

(a 8)

where $e = 2.718$. Hence

$$\sqrt{\frac{\nu}{\nu_a}} \approx 1 + \frac{2}{\pi n_0}\left(\alpha s \ln\frac{2n_0}{e n_a} - n_a\right).$$

(a 9)

There is a negative effect on ν because of cutting off the integral at n_a instead of at zero and a positive effect caused by the space charge. The first will be partially or totally offset by adding the sheath thickness. The space-charge term is already dominant if $\bar{s}$ is given its value (a 3) at n_a, but not if given the value s_0. There is a larger space charge, however, in the sheaths. The effect of this sheath charge will have to be added to the effect computed here.

(s) <u>Sheaths</u> $J = J_d$

None of the limits (c) to (a) are valid out to the boundary of the plasma. As n_0 increased and successive approximations became valid in the center, the previous
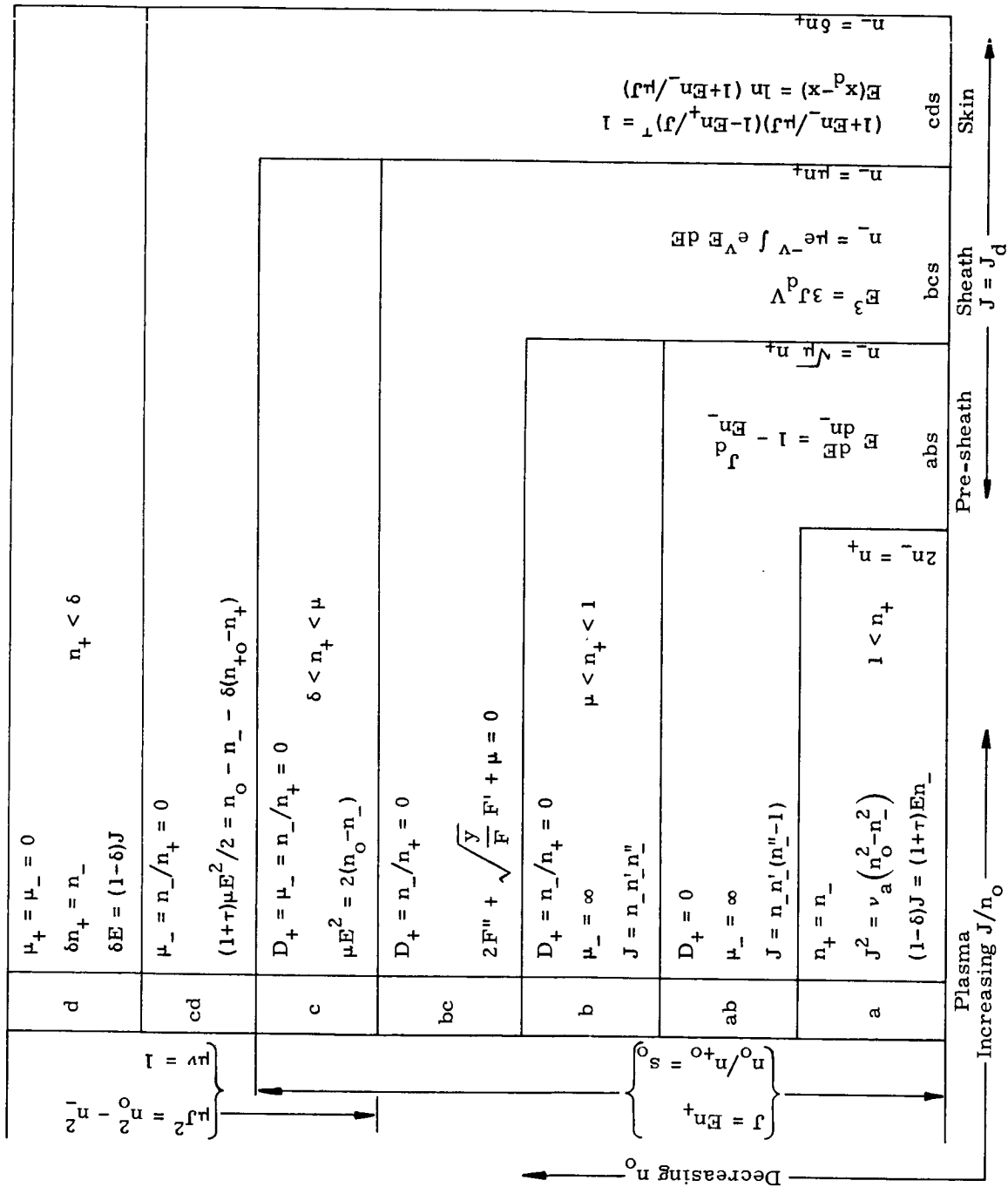
Fig. X-3. Diagrammed solutions of the ambipolar equations.

regions were merely pushed out toward the wall in successive layers. As they approach the wall, a new approximation becomes valid: the current through them becomes substantially constant. The current is therefore no longer a suitable independent variable in the equations, and we turn to the variable E by dividing each equation (e 7) by Poisson's equation:

$$\tau \frac{dn_+}{dE} = \frac{-J_d + En_+}{n_+ - n_-} \tag{s 1}$$

$$\frac{dn_-}{dE} = \frac{-\mu J_d - En_-}{n_+ - n_-} \tag{s 2}$$

$$x_s = \int \frac{dE}{n_+ - n_-}. \tag{s 3}$$

The crowding toward the wall is evident in Fig. X-1 where all the profiles crowd together at $J = J_d$. Figure X-3 is a diagram based on Fig. X-1 but distorted to show how the central regions are bounded toward the wall. Thus we see that the total sheath region may be divided into three parts: a pre-sheath, a sheath, and a skin.

(abs) <u>Pre-Sheath</u> $D_+ \to 0$ $\mu_- \to \infty$ $\sqrt{\mu} < n_-/n_+ < 1/2$

This is an extension of the space-charge transition region. It is the region that is well known to prove theorists in which electron space charge is not negligible but in which the voltage drop is sufficiently large to accelerate the ions to a drift energy corresponding to the electron temperature. C+K call it the sheath (p. 45). Its equations are

$$En_+ = J_d$$

$$E^2 \frac{dE}{dn_-} = E - J_d/n_- \tag{abs 1} (Eq. 4, p. 46)$$

This equation has to be solved numerically.

The pre-sheath is joined to the plasma where the two conditions $n_+ < 2n_-$, $J_d < 2J$ are well satisfied. This leads to inequalities

$$\frac{4}{3} n_-^2 < n_o^2 < n_-^3 \tag{abs 2}$$

which can be satisfied by

$$n_- = n_a = n_o^{4/5} \tag{abs 3}$$

but n_o had better be quite large or the overlap region (abs 2) will not be appreciable. For practical plasmas numerical solutions of (ab 2), extending to reasonably large n_o, will be required. Unfortunately there is no approximate solution from which the thickness

of the pre-sheath can be calculated.

(bcs) <u>The Sheath</u> $\tau \to 0$ $n_- \to 0$ $\mu < n_-/n_+ < \sqrt{\mu}$

In the sheath, as generally understood, the electron space charge may be neglected. We shall not, however, neglect the electron flow. The equations are then

$$En_+ = J_d$$

$$\frac{dn_-}{dE} = -\frac{E^2 n_-}{J_d} - \mu E. \qquad \text{(bcs 1)}$$

The solution of the second is

$$n_- = \mu e^{-V} \int_E^c e^V E \, dE \approx \mu n_{+c} e^{V_c - V}, \qquad \text{(bcs 2)}$$

where

$$V = E^3/3J_d \qquad \text{(Top eq., p. 48)}$$

and

$$x_c - x_b = \frac{E_c^2 - E_b^2}{2J_d} = \frac{1.04}{J_d^{1/3}} \left(V_c^{2/3} - V_b^{2/3} \right). \qquad \text{(bcs 3)}$$

The solution (bcs 2) is joined to the pre-sheath where the conditions of (abs 1) and (bcs 1) overlap, which turns out to include

$$(n_-/n_+)_b = \sqrt{\mu}. \qquad \text{(bcs 4)}$$

Thus

$$n_b^3 = n_o^3 e^{-3V_b}$$

$$n_{+b}^3 = \frac{J_d^3}{E_b^3} = \frac{J_d^2}{3V_b}$$

$$\frac{n_o}{\mu^{3/2}} = \frac{e^{3V_b}}{3V_b}. \qquad \text{(bcs 5)}$$

Here, we have taken $J_d^2 = n_o^2$ (a 2) as being close enough.
    Similarly,

$$(n_-/n_+)_c = \mu \tag{bcs 6}$$

and

$$\frac{n_o}{\mu^3} = \frac{e^{3V_c}}{3V_c}. \tag{bcs 7}$$

Equations (bcs 5) and (bcs 7) give the potentials of the sheath, and by introducing them in (bcs 3) we can evaluate the sheath thickness. In general,

$$x_c - x_b \approx \frac{V_c^{2/3}}{J_d^{1/3}} \approx \frac{\ln\left(n_o/\mu^3\right)}{3n_o^{1/3}} \tag{bcs 8}$$

but if V_b is large, so that the sheath potential $V_c - V_b$ is small,

$$x_c - x_b \approx \frac{\Delta V^{2/3}}{J_d^{1/3}} \approx \frac{\ln 1/\mu}{\left(9n_o \ln n_o/\mu^3\right)^{1/3}}. \tag{bcs 9}$$

Unfortunately the sheath thickness cannot be compared with the cutoff distance n_a/n_o in (a 8). For small values of V_b one could set $n_a = n_o^{2/3}$, and (bcs 8) shows that the sheath is thicker than the cutoff, but for large V_b there is no suitable value of n_a where the sheath and the plasma overlap. The pre-sheath thickness has to be included.

The boundary condition (bcs 6) that we have chosen corresponds to setting $\mu J = En_-$. It is therefore the point at which the Boltzmann distribution breaks down. It is because we did not go beyond this point that we could use the simpler expression in (bcs 2). It is also the point where conventional sheath theory breaks down, and the usual procedure of setting the random electron current equal to the directed ion flow is equivalent to (bcs 6). There is, however, a skin outside the sheath.

(cds) <u>The Skin</u> $J = J_d$ $E = E_d$ $\delta < n_-/n_+ < \mu$

The skin is an extension of the ion flow transition. The electron density is so small that the electron current cannot be left out of the equations, but the skin is so thin that the electric field may be taken as constant. Eliminating dE between (s 1) and (s 2), we get

$$-dx = \frac{dn_-}{\mu J + En_-} = \frac{\tau \, dn_+}{J - En_+} \tag{cds 1}$$

$$(1+En_-/\mu J)(1-En_+/J)^{\tau} = 1 \tag{cds 2} \tag{D. 1. 5}$$

$$E(x_d-x) = \ln (1+En/\mu J).$$

This has the proper limits, written

$$En_- \approx \mu J \qquad En_+ = J \qquad \qquad \text{(Eq. 2, p. 36)}$$

$$n_- \to 0 \qquad \bar{n}_- = \delta n_+ , \qquad \qquad \text{(cds 3)}$$

the thickness of the skin $\Delta x = n_-/\mu n_o$ is extremely small so that it does not affect the value of ν. For nearly all purposes, the wall can be put at x_c with the boundary condition (bcs 6).

The Characteristic Curve

Integrating the appropriate equations numerically, C+K obtain the characteristic curve shown in Fig. X-4. We show in the same figure the curve obtained by A+R. A+R
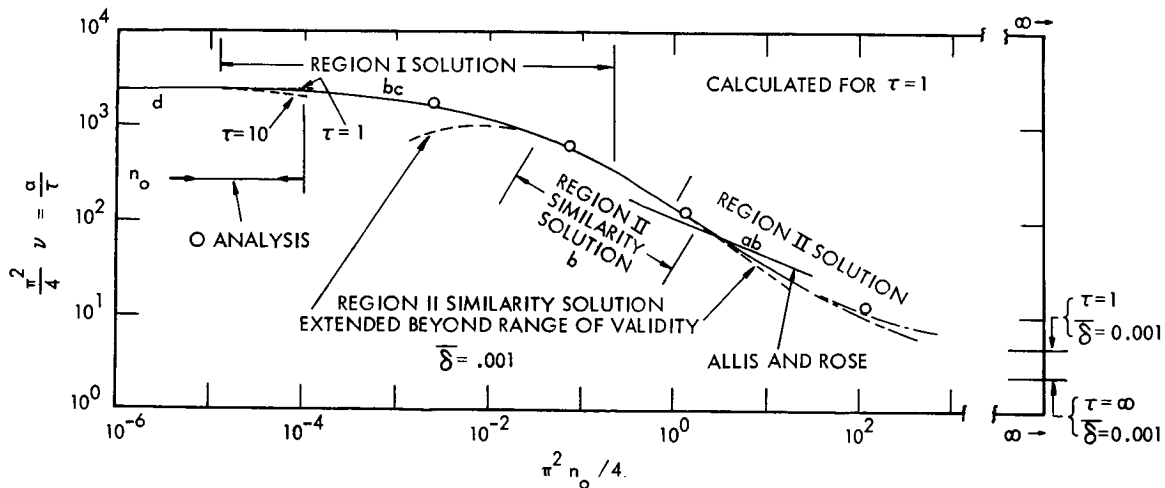
Fig. X-4. Normalized ionization frequency vs plasma density.

used $\mu = 1/32$, C+K used $\mu = 1/1000$. As 1000 is roughly the square of 32, the slope of the C+K curve is roughly twice that of A+R. There is no disagreement. The C+K curve extends farther in both directions because they have dropped the small high derivative term which is most troublesome to a computer, but it does not extend much farther, particularly in the direction of high densities which is of greatest interest. It is for this reason that A+R proposed an approximation, and that C+K in their section D. 3 study the limit $n_o \to \infty \; n_- \to 0$ and reach the conclusion $\nu \to \nu_a$ from below. As this is surprising, we must look into their analysis of this limit with some care.

(as) The Ambipolar Singularity $\qquad n_o \to \infty \qquad n_- \to 0$

In their section D. 3, C+K consider the singularity created by the somewhat contradictory limits $n_o \to \infty \; n_- \to 0$. As we have seen, the ambipolar limit is valid only for

$$n_-^3 > n_{ab} = n_0^2 \qquad \text{(as 1)}$$

but since

$$n_{ab}/n_0 = n_0^{-1/3} \to 0 \qquad \text{(eq. 4, p. 60)}$$

as

$$n_0 \to \infty,$$

the fraction of the plasma where the ambipolar limit is invalid decreases as n_0 increases.

The terms on the right of equation (e 11) are small for $n_- > n_{ab}$; hence, we can substitute from (a 2) to get

$$\frac{1 + \mu}{1 + \tau} J^2 = n_0^2 - n^2 + A n_0 \frac{n_0 - n}{n} \qquad \text{(as 2) (D. 3. 9)}$$

$$A = \frac{1 - \delta}{1 + \tau} \frac{1 - 2\tau - 3\delta}{(1+\mu)^2}.$$

We note that the new term is small of order $1/n_0$ compared with the first near the center but large of order $1/n$ near the wall. A recursive procedure will produce converging terms at the center, diverging terms at the wall. Surprisingly, C+K expand (D. 3. 9) in powers of $\xi = (x_d - x)$ about the singularity at $n = 0$ (D. 3. 10). This is dangerous because the series has no validity at its origin.

They now seek a sheath solution, making use of (e 14) which can be written

$$(1+\tau)n_- = a + \frac{E^2}{2} - \tau s + (1+\mu) \int_x^{x_d} J \, dx,$$

where $\qquad \text{(as 3) (D. 3. 11)}$

$$a = n_0 + \tau n_{+o} - (1+\mu) \int_0^{x_d} J \, dx.$$

The terms of order n cancel in "a", as can be seen by setting

$$J = \sqrt{\nu} \, n_0 \sin x/\Gamma_\nu \qquad \text{(as 4)}$$

which yields

$$a = n_0 + \tau n_{+o} - (1+\mu)\nu n_0.$$

For ν near its ambipolar value,

$$v = \frac{1+\tau}{1+\mu} + v_1/n_o \qquad\qquad \text{(as 5)}$$

$$a = \tau s_o - (1+\mu)v_1. \qquad\qquad \text{(as 6)}$$

C+K now expand (as 3) in powers of ξ, which is correct. The solution (as 3) should be joined to (as 2) in the region where both are valid, that is, for $n_- > n_a$. C+K do not do this. On pp. 68 and 69 they equate the power series term by term. This is equivalent to equating the functions and several derivatives at the origin of the series, that is, at $n = 0$; however, (as 2), their (D. 3. 9), does not even approximate the solution at $n = 0$. We have carried our approximation (as 4) farther than we should in (as 6) so that this equation can be compared with the third equation on p. 68 of C+K. Following them and neglecting τs_o, one finds that the two equations turn out to be the same but with opposite signs. Recursion procedures often give alternating signs when they diverge.

C+K find different signs for v_1 in isothermal and active plasmas. The reason for this is seen in the sign of a (a 6) in the distributed charge formula (a 9). a changes sign for $\tau = 1/(2+3\mu)$. This can be traced back to the sign of σ in (e 11) and taking the derivative of the limit formula (a 3) to insert in s'; however, s' is a sharply peaked function in the ambipolar limit and s' changes sign quickly in the pre-sheath. Thus the joining procedure is more difficult in the isothermal case and it would be better to use the first integral (e 13).

W. P. Allis

References

1. I. M. Cohen and M. D. Kruskal, Asymptotic Theory of the Positive Column of a Quiescent, Longitudinally Uniform Gas Discharge, Matt-202, Plasma Physics Laboratory, Princeton University, Princeton, N. J., July 1963.

2. W. P. Allis and D. J. Rose, The transition from free to ambipolar diffusion, Phys. Rev. 93, 84-93 (1964).

B. ION CYCLOTRON RESONANCE IN A RADIOFREQUENCY DISCHARGE

Ion cyclotron resonance has been observed in a rather unique RF discharge. As shown schematically in Fig. X-5, the discharge is produced by a single electrode that is a thermionic electron emitter (oxide-coated). The discharge appears to fill entirely a region, 10 cm in diameter by 20 cm long, formed by electric insulators. The DC magnetic field $\underset{\sim}{B}_o$ is uniform within 1 1/2 per cent over this region. Also indicated in Fig. X-5 are two magnetic pickups; the smaller is enclosed in 8-mm quartz tubing and measures the AC magnetic field that is perpendicular to $\underset{\sim}{B}_o$, the larger is wrapped on the discharge tube and measures the AC magnetic field that is parallel to $\underset{\sim}{B}_o$. Signals from these pickups are detected and then plotted directly as a function of B_o on an
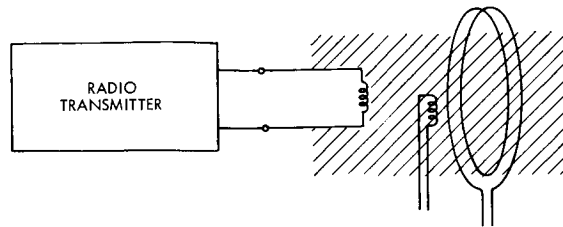
Fig. X-5. Schematic representation of the system.

x-y recorder. The frequency of the applied voltage is constant, and ion cyclotron resonance is expected when

$$\beta_+ = \frac{eB_o}{M_+ \omega} = 1.$$

Data obtained in a hydrogen discharge are shown in Figs. X-6 and X-7. The hydrogen pressure is approximately 20 μ, and the applied frequency is 4 Mc. It is assumed that the hydrogen is disassociated and thus β_+ is based on the proton mass. The curves obtained from the two pickups are roughly similar and only those for the smaller one are given here.

In Fig. X-6 the detector is tuned to 4 Mc; in Fig. X-7 it is tuned to 8 Mc. The response of a wideband detector (10 cps to 10 Mc) showed little or no resonance behavior.

The 8-Mc signal is stronger than the 4-Mc signal. In order to get comparable responses in Figs. X-6 and X-7, the 8-Mc signal was attenuated 13 db. The light showed considerable modulation at 8 Mc, little at 4 Mc.

The curves of Figs. X-6 and X-7 are labeled with the applied peak-to-peak voltage. This voltage varies little with B_o. In fact, it is essentially the same for a given setting of the transmitter controls, with or without a plasma. The shapes of the curves depend on the applied voltage. This is thought to be due to variation in the plasma density. A similar dependence on the pressure is observed.

It is tempting to relate peaks observed for $\beta_+ < 1$ to plasma resonance of the extraordinary wave propagating across $\underset{\sim}{B}_o$. The resonance condition for a lossless plasma is

$$a^2 = \frac{\left(\beta_-^2 - 1\right)\left(1 - \beta_+^2\right)}{\beta_- \beta_+ - 1} \approx \frac{M_+}{m_-}\left(1 - \beta_+^2\right),$$

where $\beta_- = e\,B_o/m_- \omega$, and the approximation applies in the vicinity of ion cyclotron resonance. A resonance occurring at $\beta_+ = 0.8$ would, then, correspond to $a = 25.7$, or a plasma density of 1.3×10^8 cm^{-3}.

It is equally tempting to relate peaks observed for $\beta_+ > 1$ to generation of ion cyclotron waves. The dispersion relation for a lossless plasma is
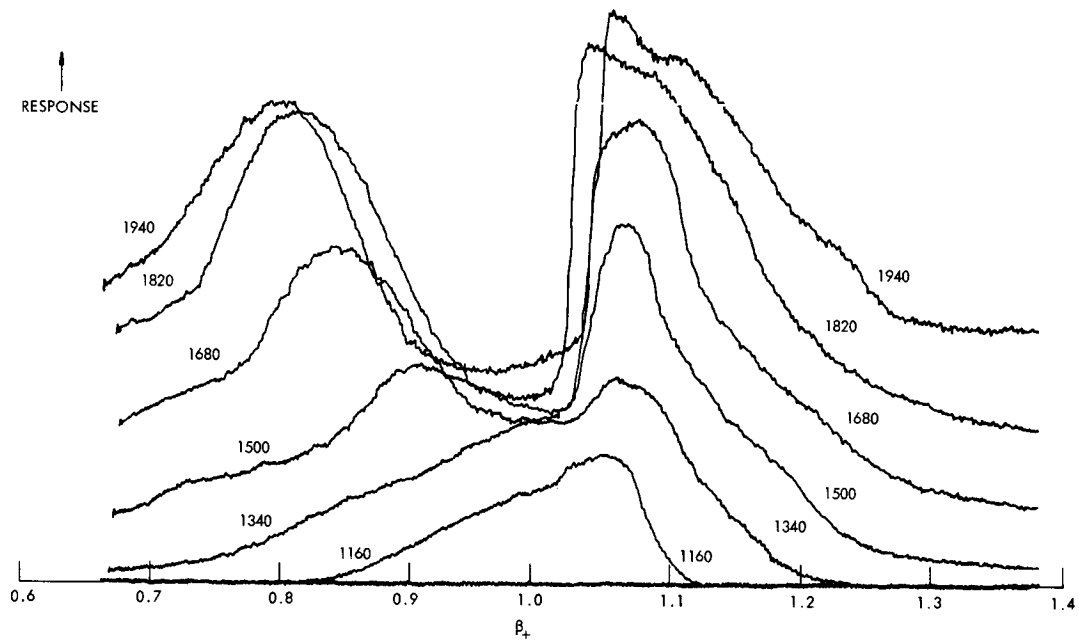
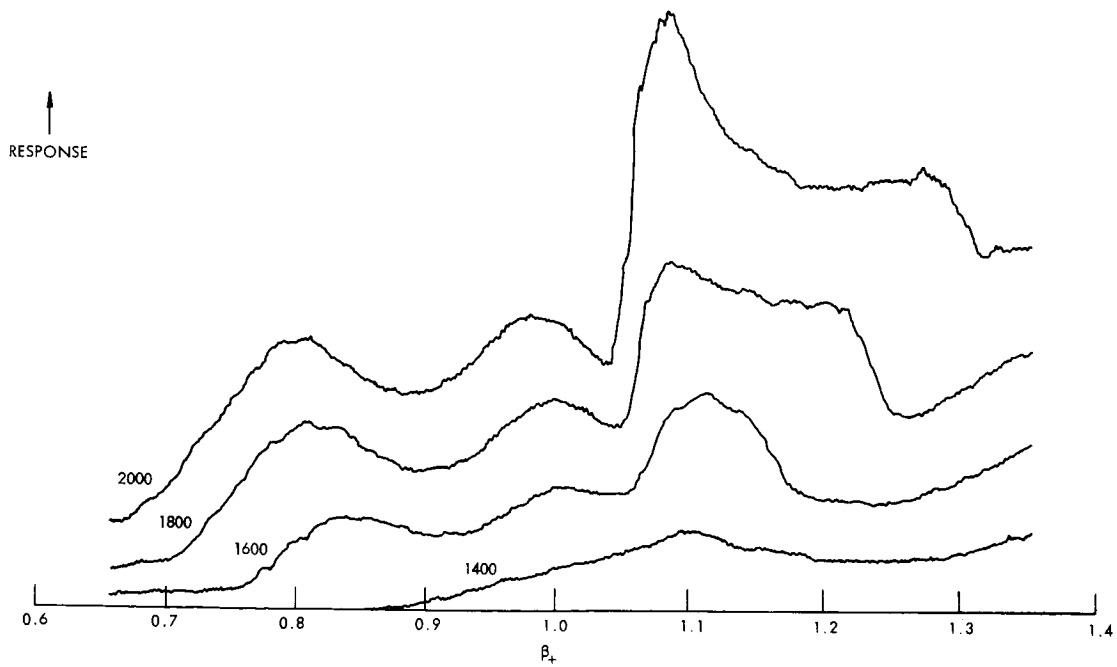Fig. X-6. Receiver response vs magnetic field, receiver tuned to 4 Mc.

Fig. X-7. Receiver response vs magnetic field, receiver tuned to 8 Mc.

$$n^2 = 1 - \frac{a^2}{(1-\beta_+)(1+\beta_-)} \approx \frac{m_-}{M_+} \frac{a^2}{\beta_+(\beta_+-1)},$$

for which the approximation again applies in the vicinity of ion cyclotron resonance. If we take the 20-cm length as approximately $\lambda/4$, $n \times 100$, since the free-space wavelength is 75 meters. The loading at $\beta_+ = 1.1$ would then correspond to $a = 1.4 \times 10^3$, or a plasma density of 4×10^{11} cm^{-3}.

If both resonances are observed the plasma resonance must occur in the outer regions of the plasma. The existence of the plasma resonance has not yet been verified. Distinct peaks, however, have occasionally been observed at $\beta_+ = 1$ and $\beta_+ > 1$. The second peak then grows, relative to the first, and moves to higher β_+ as the applied voltage or pressure is increased. This gives some support to the ion cyclotron wave interpretation.

<div align="right">J. J. Nolan, Jr.</div>

C. ELECTRON CYCLOTRON ABSORPTION IN THE CESIUM AFTERGLOW

A knowledge of the electron-atom collision frequency in cesium as a function of electron velocity is essential in order to calculate transport coefficients in the thermionic energy converter, which uses cesium plasma as the conducting medium.

Many experimental measurements[1] have been made of this quantity, no two agreeing entirely. The disagreement arises from the limited range of electron energies over which the data were valid, the uncertainty of the electron energy, and the differences in the techniques of measurement. Recently Stone and Reitz[2] have given a theory for the collision cross section of slow electrons with cesium atoms. The experimental data is rather widely scattered about their theoretical curve.

In order to obtain a set of data connecting the limited regions of experimentation of other workers, an experiment has been devised in which the electron-atom collision frequency can be measured over a wider range of electron energies. Preliminary measurements have been made. The wide range of energies is obtained from the afterglow period of a pulsed DC discharge in cesium. During this period the electron temperature decays from about 5000°K to 500°K. By using the transient microwave radiometer,[3] it is possible to sample radiation emitted by and transmitted through the plasma during a short interval of time (1, 20, 200 or 1000 µsec). The choice of a certain time in the afterglow at which to sample fixes an electron temperature, which may be measured by using the radiometer technique. Then by employing a highly attenuated klystron probing signal of magnitude only one hundred times that of the thermal radiation from the plasma (so that the probing signal does not heat the plasma), the transmission of the plasma is measured at this temperature with the radiometer now used as a detector for the klystron

signal. The klystron signal is propagated through the plasma in a direction parallel to a DC magnetic field that is applied to the plasma in such a way that when the signal electric field is of the proper frequency it will interact strongly with the plasma electrons as they orbit in the magnetic field. The transmission of the plasma is then measured as a function of the magnetic field and the width of this cyclotron resonance can be related
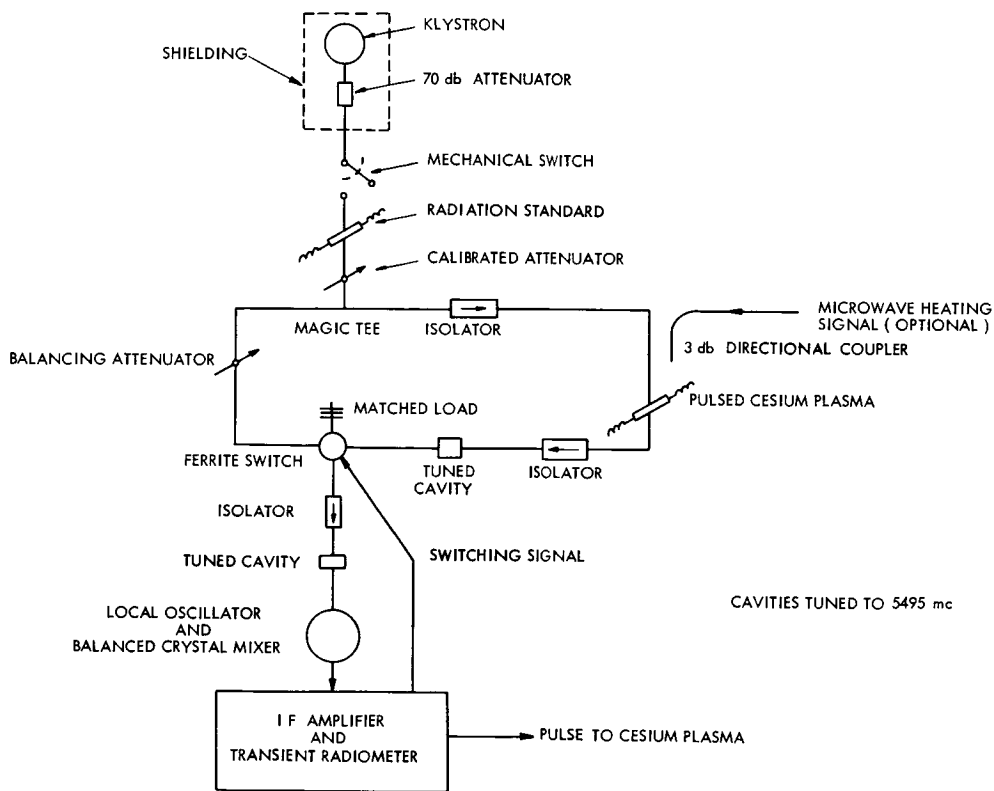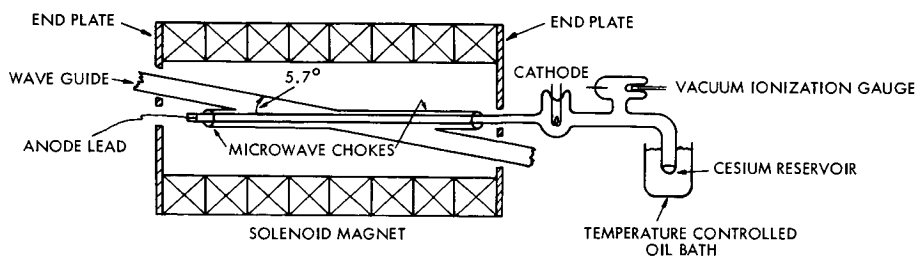

Fig. X-8. Microwave circuit for transient absorption and temperature measurements.

Fig. X-9. Cesium quartz discharge tube and magnetic field configuration.

to the electron-atom collisions which interrupt the motion of the orbiting electrons.

1. Experimental Arrangement

Figure X-8 shows the microwave bridge circuit that is used to measure transient temperatures and plasma absorption. The details of the temperature measurement have been described in previous reports.[3] When absorption measurements are to be made, the radiation standard is turned off, the klystron at the top of the figure is switched on, and the plasma absorption in the right arm of the bridge is balanced with the attenuator in the left arm.

The cesium plasma is contained in a 1.3 × 75 cm cylindrical quartz discharge tube that is inserted at an angle of 5.7° through the broad face of a 1 × 2 inch rectangular brass waveguide (See Fig. X-9). Microwave chokes prevent radiation leakage from the waveguide. A pulsed DC discharge is used to produce the plasma; the currents and voltages are about 20 ma and 200 v at a cesium pressure of 0.03 torr. The cesium pressure is determined by the vapor pressure of cesium vapor in equilibrium with a cesium puddle in a reservoir whose temperature is carefully controlled by an oil bath. The rest of the tube is maintained ~50°C hotter than the reservoir by using electrothermal heating tapes and thermocouples to monitor the temperature.

2. Theory

The absorption coefficient of a tenuous, waveguide-contained plasma that is longitudinally magnetized may be calculated by using a perturbation theory in which the unperturbed electromagnetic field is given by the modes of the empty waveguide. The general equations have been given by Bers[4]; the result of the perturbation calculation for a, the power absorption coefficient, is

$$
a = -\frac{\left(\frac{1}{10}\right)\omega_{p_0}^2}{2c\sqrt{1-\left(\frac{\lambda}{2a}\right)^2}}\left(-\frac{4\pi}{3}\int_0^\infty \left(\frac{\nu_c}{(\omega-\omega_b)^2+\nu_c^2}+\frac{\nu_c}{(\omega+\omega_b)^2+\nu_c^2}\right)\frac{df^0}{dv}\,v^3\,dv\right),
$$
(1)

where

$$\omega_{p_0}^2 = \frac{n_0 e^2}{m\epsilon_0},$$ the square of the electron plasma frequency

c = velocity of light

$$\omega_b = \frac{eB}{m},$$ the electron cyclotron frequency

ν_c = electron-atom collision frequency

$\omega = \dfrac{2\pi c}{\lambda}$, radian frequency of the probing wave

v = electron speed

f^o = electron velocity distribution in absence of the probing signal

n_o = electron density at the center of the discharge tube

a = the wider dimension of the rectangular waveguide.

The factor of 1/10 corrects for the fact that the plasma fills only the center portion of the waveguide cross section. The electron density is assumed to be distributed in the fundamental diffusion mode across the discharge tube cross section, and the magnetic field is assumed parallel to the waveguide axis.

The perturbation calculation is only valid provided the plasma index of refraction does not deviate significantly from one.

A more detailed calculation of a and of the plasma emission coefficient is in progress. When the plasma electrons have a Maxwellian distribution of velocities $\dfrac{df^o}{dv} = -\dfrac{mv}{kT}f^o$, and the magnetic fields are near cyclotron resonance, a may be written as

$$a = -\frac{\left(\frac{1}{10}\right)\omega_{p_o}^2}{2c\sqrt{1-\left(\frac{\lambda}{2a}\right)^2}}\left(\frac{4\pi m}{3kT}\int_0^\infty \frac{v^4 f^o v_c}{v_c^2 + (\omega-\omega_b)^2}dv\right) \tag{2}$$
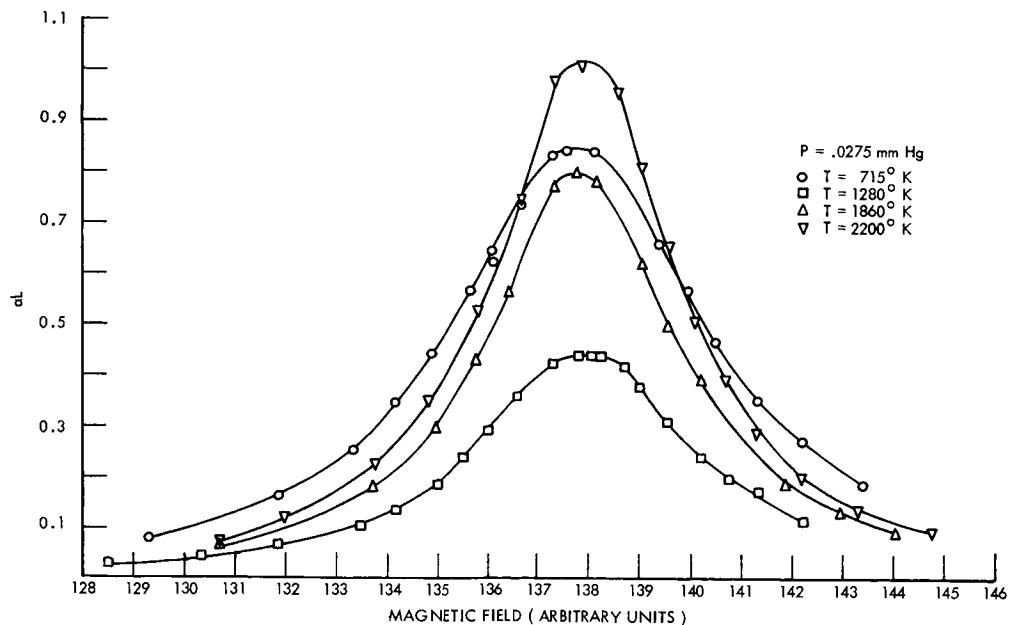
Fig. X-10. Dependence of plasma absorption (aL) on magnetic field for various electron temperatures at a pressure of 0.0275 torr.
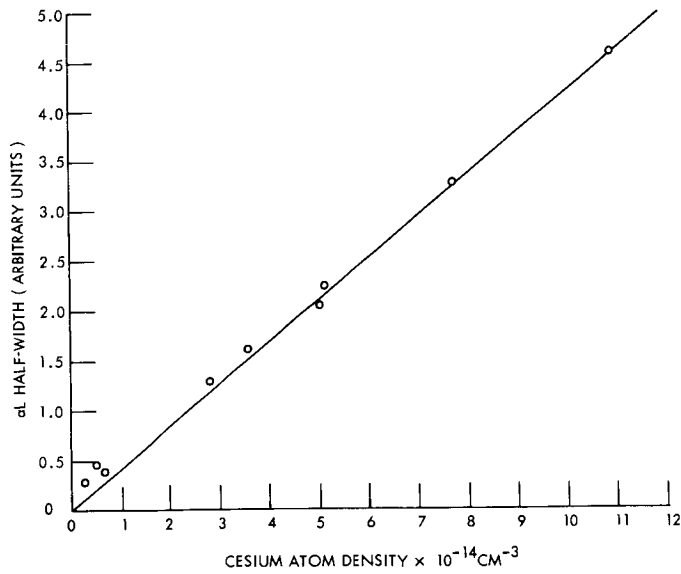
Fig. X-11. Half-width of αL versus cesium atom density.
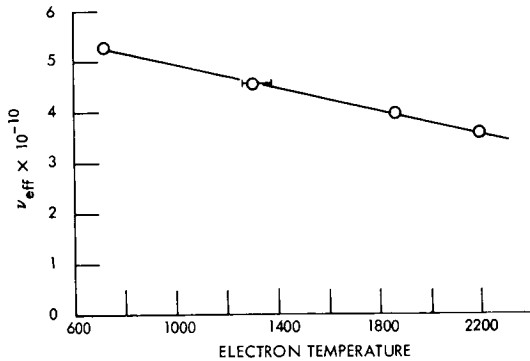
Fig. X-12. Effective collision frequency, calculated from half-widths of Fig. X-10, versus electron temperature.

+   PRESENT DATA
×   R.B. BRODE
△   C. BOECKNER AND F.L. MOHLER
◨   G.J. MULLANEY AND N.R. DIBELIUS
▼   N.D. MORGULIS AND YU P. KORCHEVOI
●   C.L. CHEN AND M. RAETHER
▲   D.N. MIRLIN, G.E. PIKUS AND V.G. YUR'EV
▽   D. ROEHLING
○   R.K. FLAVIN AND R.G. MEYERAND, JR.

Fig. X-13. P_c for cesium, determined by various experimenters.

The experimental study is based on the measurement of a as a function of ω_b. From a knowledge of the Maxwellian nature of f^0, the measured electron temperature, and the shape of the resonance of a, it is possible to determine the velocity dependence of ν_c.

3. Experimental Results

Preliminary measurements of the dependence of a on the magnetic field have been made for several different cesium atom densities corresponding to pressures up to 0.059 torr at an electron temperature of 2200°K; similar measurements have been made for various electron temperatures at a pressure of 0.0275 torr.

Figure X-10 gives the data taken at fixed pressure and shows aL ($L \approx 20$ cm is the effective length of the discharge tube in the waveguide) as a function of magnetic field. The change in the shape of the resonance with temperature reflects a change in the velocity dependence of ν_c. Figure X-11 shows the results at fixed temperature, and the half-width of aL is plotted versus cesium atom density. At higher pressures the half-width is linear with density. The nonlinearity at low densities is caused by magnetic field inhomogeneities and by Doppler broadening resulting from electron thermal motion along the magnetic field lines.

An effective collision frequency ν_{eff} may be calculated from the half-width expressed in units of frequency and plotted against electron temperature to gain a qualita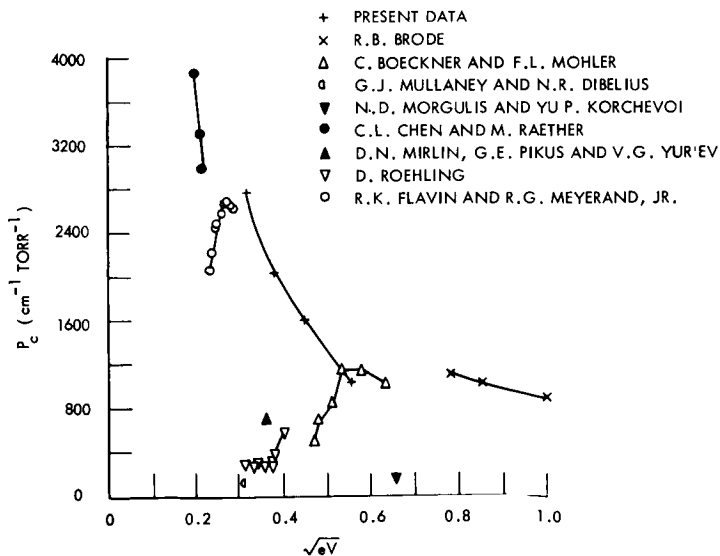tive feeling for the energy dependence of the real collision frequency. This effective collision frequency, evaluated for a density of cesium atoms corresponding to 1 torr and 273°K, is plotted in Fig. X-12 and shows a tendency to decrease with increasing temperature.

To make a satisfactory comparison with other experiments, the correct velocity dependence of ν_c must be determined by assuming a polynomial velocity dependence of ν_c, fitting the resonance curves of Fig. X-10, and then calculating the collision probability $P_c = \nu_c/v$. We can, however, calculate $P_{eff} = \nu_{eff}/(\overline{v^2})^{1/2}$ and compare this with other experimental data. $P_{eff} = P_c$ only when ν_c is independent of electron velocity. The results of this approximation are plotted along with the experimental results of others in Fig. X-13. If a more quantitative analysis of the data were made, stronger conclusions might be drawn. At present, the great variation in the experimental data serves as a stimulus to further work.

Three other points are worth mentioning at this time. For all measurements of this experiment the electron velocity distribution was ascertained to be Maxwellian by checking that the measured electron temperature showed no resonance at cyclotron frequency.[5] This required operating at peak values of aL as large as one (see Fig. X-10). The condition that the plasma index of refraction be ~1 for all values of ω_b is satisfied provided that $\dfrac{\omega_p^2}{2\omega_b\nu_c} \ll 1$. For a velocity-independent ν_c, if we use Eq. 2, $(aL)_{max} \approx \dfrac{L}{20c}\dfrac{\omega_p^2}{\nu_c}$,

so that the condition may be written as

$$\frac{\omega_p^2}{2\omega_b \nu_c} = \frac{(\alpha L)_{max} 20c}{L\, 2\omega_b} = (\alpha L)_{max}\left(\frac{10\lambda}{2\pi L}\right) \ll 1. \tag{3}$$

But λ = 6 cm and L = 20 so that for $(\alpha L)_{max}$ = 1, this quantity is ~1/2. In spite of this, no change in the symmetry of the lines was found for the large values of αL.

The third point is that a microwave heating field will be employed to widen the spectrum of available electron temperatures in the afterglow. The experimental provision made for this is shown in Fig. X-8.

<div align="right">J. C. Ingraham</div>

<div align="center">References</div>

1. R. B. Brode, Phys. Rev. 34, 673 (1929); C. Boeckner and F. L. Mohler, Bur. Std. J. Res. 10, 357 (1933); G. J. Mullaney and N. R. Dibelius, Am. Rocket Soc. J. 31, 1575 (1961); N. D. Morgulis and Yu. P. Korchevoi, Zh. Tekhn. Fiz. (English transl.: Soviet Phys. — Tech. Phys.), Vol. 32, No. 7, pp. 900-902 (1962); C. L. Chen and M. Raether, Phys. Rev. 128, 2679 (1962); D. N. Mirlin, G. E. Pikus, and V. G. Yurév, Soviet Phys. — Tech. Phys. 7, 559 (1962); D. Roehling, Symposium on Thermionic Power Conversion, Los Alamos Scientific Laboratory, May 1962; R. K. Flavin and R. G. Meyerand, Advanced Energy Conversion (Pergamon Press, New York, 1963), Vol. 3, pp. 3-18.

2. P. M. Stone and J. R. Reitz, Elastic scattering of slow electrons by cesium atoms, Phys. Rev. 131, 2101 (1963).

3. J. C. Ingraham and J. J. McCarthy, Quarterly Progress Report No. 64, Research Laboratory of Electronics, M.I.T., January 15, 1962, p. 76; J. C. Ingraham, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, p. 83.

4. A. Bers in W. P. Allis, S. J. Buchsbaum, and A. Bers, Waves in Anisotropic Plasmas, (The M.I.T. Press, Cambridge, Mass., 1963), p. 136.

5. H. Fields, G. Bekefi, and S. C. Brown, Phys. Rev. 129, 506 (1963).

D. RADIOFREQUENCY CYCLOTRON CONFINEMENT

Work has been extended on the theory of RF confinement at exact ion cyclotron resonance, discussed by Whitehouse,[1] Kulínski,[2] and others. Their theories do not explicitly treat the way the plasma modifies the fields. The present work has made it possible to estimate the penetration of an RF field into a plasma in the geometry discussed by Whitehouse, and then to compute the modifications to the RF confinement.

A two-dimensional geometry is used to make the problem tractable. (See Fig. X-14.) The conductor is excited antisymmetrically by the surface current J, which produces a
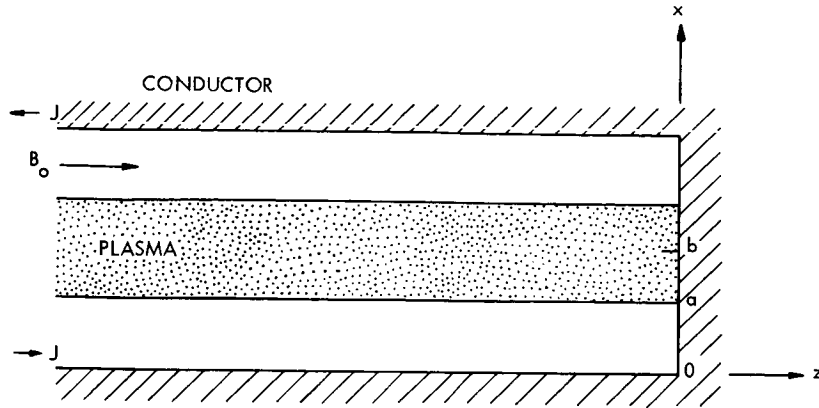
Fig. X-14. Cross section of plasma slab and conductor.

magnetic field H_y out of the paper. The problem is to determine the penetration of the magnetic field into the plasma whose dielectric constant may be extremely large. We assume a diagonal form of the dielectric tensor but allow $K_\perp$ to be different from $K_\parallel$. This problem would be very combersome if handled in terms of the standard E_z modes but becomes very simple if handled in terms of H_y modes. Using Maxwell's equations we then arrive at the following equations

$$-\frac{\partial H_y}{\partial z} = i\omega\epsilon_o K_\perp E_x \tag{1a}$$

$$\frac{\partial H_y}{\partial x} = i\omega\epsilon_o K_\parallel E_z \tag{1b}$$

$$\frac{1}{K_\perp}\frac{\partial^2 H_y}{\partial z^2} + \frac{1}{K_\parallel}\frac{\partial^2 H_y}{\partial x^2} + \frac{\omega^2}{c^2} H_y = 0, \tag{2}$$

where the first two will be used to find E_x and E_z once H_y is known. We assume an H_y of the form

$$H_y = H\, e^{ikz+i\omega t} \times \begin{cases} e^{i\beta x} & \text{inside the plasma} \\ e^{-\delta x} & \text{outside the plasma} \end{cases}$$

Using the boundary conditions and the antisymmetric excitation we arrive at the determinantal equation.

$$\delta \tanh \delta a = \frac{\beta}{K_\parallel} \tan \beta(b-a) \tag{3}$$

This equation determines k since both δ and β are functions of k and ω by (2). If we now assume that $\beta(b-a) \ll \frac{\pi}{2}$, we can expand the tangents on either side of (3) and arrive at an approximate equation for k,

$$k \approx \frac{\omega}{c} \frac{b}{a}. \tag{4}$$

The condition for validity of Eq. 4 can be expressed in another form

$$\frac{\omega_p}{c} (b-a) \ll \frac{\pi}{2}. \tag{5}$$

This approximate value for k will be useful for interpreting the general field solution inside the plasma. For $b \leqslant x \leqslant a$,

$$H_y = J \cosh \delta a \, [\cos \beta(x-a) + \tan \beta(b-a) \sin \beta(x-a)] \cos kz \tag{6a}$$

$$E_z = -\frac{\beta}{K_{||}} \frac{J \cosh \delta a}{i\omega\epsilon_o} [\sin \beta(x-a) - \tan \beta(b-a) \cos \beta(x-a)] \cos kz \tag{6b}$$

$$E_x = \frac{kj \cosh \delta a}{i\omega\epsilon_o K_\perp} [\cos \beta(x-a) + \tan \beta(b-a) \sin \beta(x-a)] \sin kz \tag{6c}$$

It is clear that H_y and E_x are symmetric about the line $x = b$, while E_z is antisymmetric about this line. These equations coupled with the approximate determinantal equation now answer the question whether or not H_y penetrates into the plasma slab. Even though a large $K_\perp$ may shield the plasma from E_x, the magnetic field H_y will penetrate if $\beta(b-a)$ is considerably less than $\frac{\pi}{2}$. From (6b) it can be seen that β is associated with an E_z field. Since this field is parallel to the plasma boundary, it cannot be shielded out by the plasma. It is the E_z field that allows the magnetic field to penetrate the plasma.

Returning now to the problem of the particle's orbit treated by Whitehouse, we use Eq. 6a and 6c in the limit $\beta(b-a) \to 0$ to specify the H_y and E_x. This assumes a thin slab of plasma. Using these fields we derive the following equation for the z velocity of a particle

$$\dddot{z} + \left[\Omega_\perp^2 + (\omega \mp \Omega_z)^2\right] \dot{z} + \frac{\omega(\omega \mp \Omega_z) \, \Omega_\perp^2}{K_\perp} z = 0, \tag{7}$$

which differs from Whitehouse's Eq. 7 only by the presence of $K_\perp$. At resonance the equation for the z motion is the same, giving identical z reflection. However, the transverse energy that a reflected particle picks up during reflection will be quite different from that computed by Kulinski for reflection in a plane wave. He finds[3] that the ratio of the transverse energy, at the point where the particle's Z motion just stops,

to its initial z energy is approximately $4\left(\dfrac{B_o}{B_{RF}}\right)^2$. For typical values of B_o and B_{RF}, this ratio can be extremely large and puts a severe restriction on the design of a simple experiment. Our theory indicates that E_x will be approximately $\dfrac{1}{K_\perp}$ times the free space E_x' that Kulinski considers, and therefore a particle can be reflected without an extreme increase in its transverse energy.

P. W. Jameson

References

1. Quarterly Progress Report No. 73, Research Laboratory of Electronics, M.I.T., April 15, 1964, p. 40.

2. S. Kulinski, Rijnhuizen Report 63-14, FOM-Instituut Voor Plasma-Fysica, Jutphass, Netherlands, November, 1963.

3. S. Kulinski, Ibid., p. 19, Eq. 66.

E. DEFINITION OF THE "RAY REFRACTIVE INDEX" AND ITS ROLE IN THE RADIATION IN ANISOTROPIC PLASMAS

Studies of the flow of radiation in inhomogeneous, dispersive media are based on a classical theorem of geometrical optics by Clausius. This theorem can be stated as follows: Consider two points O_1 and O_2 of the medium (Fig. X-15), lying within elementary areas da_1 and da_2, respectively. Let R be the distance between O_1 and O_2 and let the
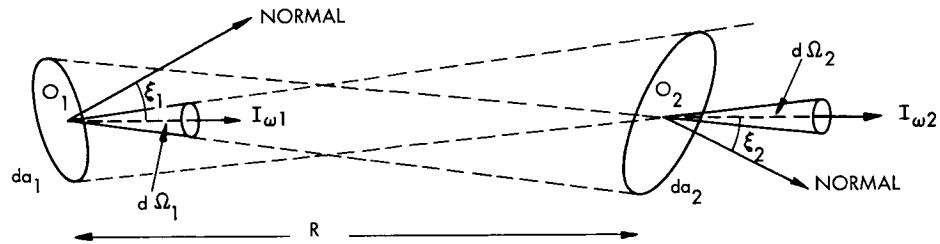
Fig. X-15. Rays between two points, O_1 and O_2, at which the respective intensities of radiation are $I_{\omega 1}$ and $I_{\omega 2}$.

normal to da_1 make an angle ξ_1 with O_1O_2, and the normal to da_2 make an angle ξ_2 with O_1O_2. The distance R is taken to be large compared with da_1 and da_2 so that all pencils of radiation passing through a point of da_1 and filling da_2 have essentially the same solid angle $d\Omega_1$; similarly, $d\Omega_2$ is the solid angle of every pencil of rays passing da_2 and filling area da_1. Then if n_1 and n_2 are the refractive indices at O_1 and O_2, as a consequence of Snell's law of refraction,

$$n_1^2 \cos \xi_1 \, da_1 d\Omega_1 = n_2^2 \cos \xi_2 \, da_2 d\Omega_2,$$

or

$$n^2 \cos \xi \, da d\Omega = \text{constant along the ray.} \tag{1}$$

This is Clausius's theorem. The theorem is not valid, however, in anisotropic dispersive media. Nevertheless, one may write

$$n_r^2 \cos \xi \, da d\Omega = \text{constant along the ray.} \tag{2}$$

Here, the quantity n_r plays the role of n in isotropic materials. For lack of better terminology, we call n_r the "ray refractive index."

It is found that n_r is related to n and to the magnitude of the group velocity w through

$$n_r^2 = \left| n^2 \frac{w}{c} \frac{d\Omega_k}{d\Omega} \left(\frac{\partial \omega n}{\partial \omega} \right)_{\theta, \phi} \right|, \tag{3}$$

MAGNETIC FIELD
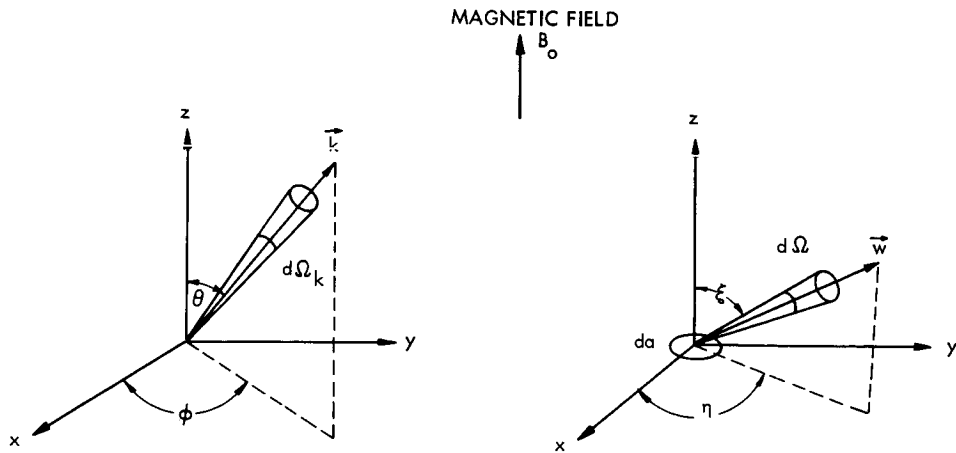
Fig. X-16. Direction of wave vector $\vec{k}$ and of group velocity vector $\vec{w}$, at a point in the medium. The elementary area, da, across which the flux of radiation is considered, is taken to lie in the xy plane.

where $d\Omega = \sin \xi \, d\xi d\eta$ is the elementary solid angle about the ray direction, and $d\Omega_k = \sin \theta \, d\theta d\phi$ is the elementary solid angle about the wave-vector direction with which the bundle of rays is associated. (See Fig. X-16.) There is a set of equations like (2) and (3) at a point r of the medium for every independent solution (that is, characteristic mode) of the dispersion equation

$$\omega = \omega(\vec{k}, \vec{r}).$$

(4)

Note that n_r is zero for all ray directions associated with angles θ and ϕ for which $k(\theta,\phi)$ is imaginary.

In the case of a plasma immersed in a magnetic field B_o, n is independent of the azimuthal angle ϕ, if the direction of B_o is chosen as the z axis of the polar coordinates z, θ, ϕ. One then finds that

$$\tan (\xi-\theta) = -\frac{1}{n}\left(\frac{\partial n}{\partial \theta}\right)_\omega$$

$$w = \frac{c}{\left(\frac{\partial \omega n}{\partial \omega}\right)_\theta} \left[1 + \left(\frac{1}{n}\frac{\partial n}{\partial \theta}\right)^2_\omega\right]^{1/2}$$

(5)

and

$$d\Omega = -\frac{d\Omega_k}{\sin \theta} \frac{\partial}{\partial \theta}\left\{\frac{n \cos \theta + \left(\frac{\partial n}{\partial \theta}\right)_\omega \sin \theta}{\left[n^2 + \left(\frac{\partial n}{\partial \theta}\right)^2_\omega\right]^{1/2}}\right\}$$

(6)

with the result that (3) becomes

$$n_r^2 = \left| n^2 \sin\theta \; \frac{\left[1 + \left(\frac{1}{n}\frac{\partial n}{\partial\theta}\right)_\omega^2\right]^{1/2}}{\frac{\partial}{\partial\theta}\left\{\frac{\cos\theta + \left(\frac{1}{n}\frac{\partial n}{\partial\theta}\right)_\omega \sin\theta}{\left[1 + \left(\frac{1}{n}\frac{\partial n}{\partial\theta}\right)_\omega^2\right]^{1/2}}\right\}} \right. \tag{7}$$

This formula holds both for electromagnetic and longitudinal (electrostatic) waves. When the medium is isotropic, n_r reduces to n.

The "ray refractive index" appears, often in disguised form, in the work of many authors[1-4] (in particular, Rytov and Mercier) who have engaged in the study of thermal and nonthermal radiation in plasmas and other anisotropic media. Before examining n_r in more detail, we summarize some of the more important results.

The results are all based on the following assumptions. The scale length over which n varies appreciably must be large compared with the wavelength in the medium. Thus (i) waves can be considered locally as plane waves satisfying the dispersion equation (4); (ii) the characteristic modes of propagation are not coupled to one another; and (iii) there are no waves reflected at gradients of n.

If there is any damping of the waves in the medium (k complex), then this damping must be very weak. This allows one to neglect damping in calculating ray trajectories and in deriving equations like Eqs. 2-7. It also allows one to define unambiguously quantities like group velocity and energy density of radiation.

1. Equations of Transport of Radiation

In a perfect dielectric which does not emit, absorb, or scatter radiation, the specific intensity I_ω obeys the equation

$$\frac{I_\omega}{n_r^2} = \text{constant along ray,} \tag{8}$$

where I_ω is the energy crossing a unit area per second per unit solid angle per unit radian frequency interval for the characteristic mode in question. Equation 8 is derived from the equations of energy conservation and by application of Eq. 2.

For a medium which both emits and absorbs,

$$n_r^2 \frac{d}{ds}\left(\frac{I_\omega}{n_r^2}\right) = j_\omega - a_\omega I_\omega, \tag{9}$$

where j_ω and a_ω are the emission and absorption coefficients, respectively, and ds is an element of length along the ray direction. The coefficient a_ω can be determined from the dispersion equation (4) and from Eq. 5 through

$$\alpha_\omega = (-2 \text{ Im } k) \cos (\xi - \theta). \tag{10}$$

2. Black-Body Radiation

a. Energy Density

The spectral energy density of black-body radiation for each characteristic mode in an anisotropic medium is given by

$$u_\omega = \frac{u_{o\omega}}{4\pi} \int_{\Omega_k} n^2 \left| \frac{\partial \omega n}{\partial \omega} \right|_\theta d\Omega_k$$

$$= \frac{u_{o\omega}}{4\pi} \int_\Omega \frac{c}{w} n_r^2 d\Omega, \tag{11}$$

where $u_{o\omega}$, the energy density in vacuum (for one polarization), is

$$u_{o\omega} = \frac{\hbar \omega^3}{2\pi^2 c^3} [e^{\hbar\omega/\kappa T} - 1]^{-1}, \tag{12}$$

b. Intensity

Since in general u_ω and I_ω are related through

$$u_\omega = \int_\Omega \frac{I_\omega}{w} d\Omega, \tag{13}$$

it follows from (3), (5), (6), and (11) that the black-body intensity, denoted by B_ω, is

$$B_\omega = n_r^2 B_{o\omega}, \tag{14}$$

where $B_{o\omega}$, the black-body intensity in vacuum (for one polarization), is

$$B_{o\omega} = \frac{\hbar \omega^3}{8\pi^3 c^2} [e^{\hbar\omega/\kappa T} - 1]^{-1}. \tag{15}$$

c. Kirchhoff's Law

In thermal equilibrium, Kirchhoff's law relates j_ω, α_ω, and $B_{o\omega}$ through

$$\frac{j_\omega}{\alpha_\omega} = n_r^2 B_{o\omega}. \tag{16}$$

When thermal equilibrium does not exist, j_ω and α_ω may be determined separately in terms of the particle momentum distribution functions $f(\vec{p})$ and $f(\vec{p}')$, where $\vec{p}$ and $\vec{p}'$ refer to the two states that participate in the emission of a quantum $\hbar\omega$. Thus

$$j_\omega = \int \eta_\omega(\vec{p}') \, f(\vec{p}') \, d^3p'$$

(17)

$$a_\omega = \int \eta_{\omega A}(\vec{p}) \, f(\vec{p}) \, d^3p - \int \eta_{\omega S}(\vec{p}') \, f(\vec{p}') \, d^3p',$$

where $\eta_\omega, \eta_{\omega A}$ and $\eta_{\omega S}$ are the differential coefficients for spontaneous emission, absorption, and stimulated emission, respectively, and are related (much as Einstein's A and B coefficients) through

$$\eta_\omega(\vec{p}') = n_r^2 \, \frac{\hbar\omega^3}{8\pi^3 c^2} \, \eta_{\omega S}(\vec{p}')$$

(18)

$$\eta_{\omega A}(\vec{p}) \, d^3p = \eta_{\omega S}(\vec{p}') \, d^3p'.$$

Individually, these coefficients have the property that

$$\eta_\omega \propto \frac{c}{w} \, n_r^2$$

$$\eta_{\omega A} \propto \frac{c}{w}$$

(19)

$$\eta_{\omega S} \propto \frac{c}{w}.$$

3. Polar Diagrams of $(1/n)$ and $(1/n_r)$

Several geometrical representations are used in describing wave propagation in an anisotropic medium. One such representation is in terms of the "phase velocity surface," which is a polar plot of the phase velocity

$$\vec{v} = \frac{\omega}{k^2} \, \vec{k} = c \, \frac{\vec{n}}{n^2}$$

(20)

as a function of the angle θ. From Eq. 20, this is equivalent (except for a constant of proportionality) to a polar plot of $(1/n)$. Much use of such diagrams has been made by Allis, Buchsbaum, and Bers[5] in classifying waves in anisotropic plasmas. In the spirit of this classification one may also inquire into the behavior of the ray refractive index and thus obtain information about the angular distribution of radiant energy. (See, for example, Eqs. 8 and 14.) Figure X-17 shows polar plots of $(1/n)$, which are the same as those of Allis, and polar plots of $(1/n_r)$. The results were derived from the computations by Rytov.[1] They apply to the case of a cold plasma, (thermal motion of electrons is neglected) with stationary ions.

Figure X-18 shows the location of the polar diagrams in relation to a Cartesian plot of the variables $\beta^2 = (\omega_c/\omega)^2$ versus $\alpha^2 = (\omega_p/\omega)^2$, where ω_c and ω_p are the electron
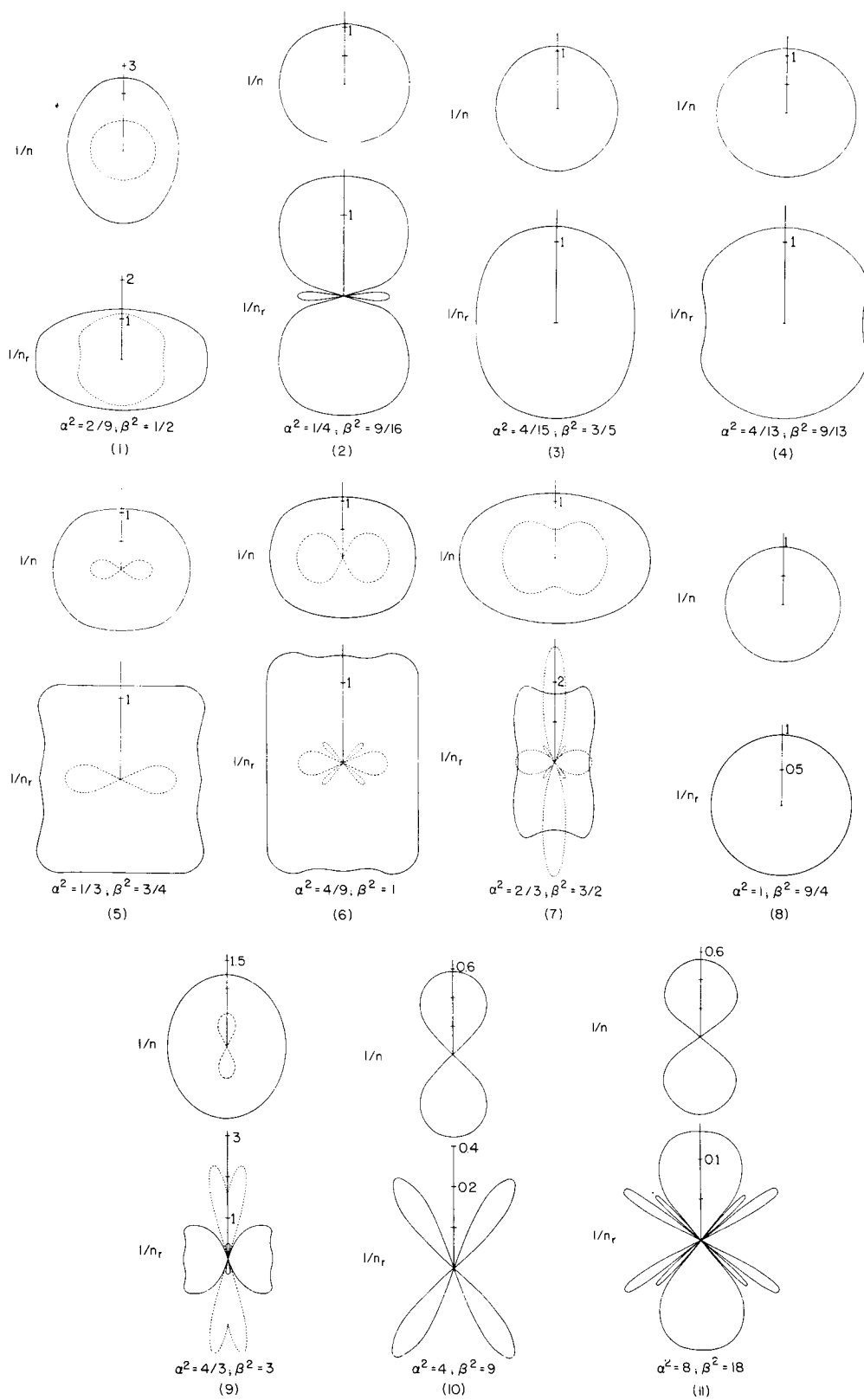
Fig. X-17. Polar plots of $1/n(\theta)$ and $1/n_r(\theta)$. The magnetic field points up ($\uparrow$) the page. The plots are to scale. The diagrams having solid lines refer to the faster of the two electromagnetic waves. Numbering of the figures (1 to 11) corresponds to the numbering in Fig. X-18. In (8), n_r is the same for both waves; but n is zero for one and unity for the other wave.

cyclotron and plasma frequencies, respectively. We note that all polar diagrams of Fig. X-17 lie on a diagonal passing through the origin of Fig. X-18. This describes a
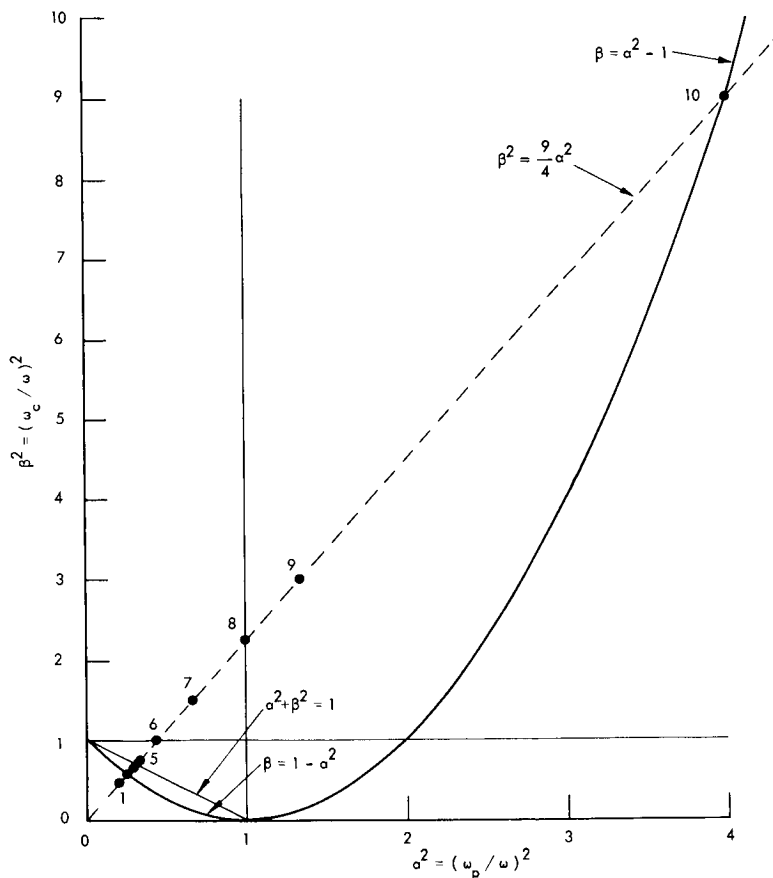
Fig. X-18. The points numbered 1, 2, 3 ... 10 refer to the location in the α^2-β^2 plane of the polar plots shown in Fig. X-17. Polar plot 11 is outside the range of this figure. The significance of all the other straight lines and curves has been discussed by Allis, Buchsbaum, and Bers.[5]

physical situation in which the plasma properties ω_p and ω_c are held fixed and the frequency of observation ω is varied from low frequencies (upper right) to high frequencies (lower left) of Fig. X-18.

For each value of (α^2, β^2) there are two polar plots of $(1/n)$ and $(1/n_r)$. They refer to the two independent electromagnetic waves, called the ordinary and extraordinary in ionospheric studies, and denoted below by superscripts (1) and (2).

For a given direction of wave propagation, the magnitude of n_r^2 is proportional to the radiation intensity. Also, the relative magnitudes of n_r^2 for the two waves is a measure of the degree of polarization of the radiation. Define

$$\text{Degree of polarization} = \frac{I_\omega^{(1)}(\theta) - I_\omega^{(2)}(\theta)}{I_\omega^{(1)}(\theta) + I_\omega^{(2)}(\theta)} \; ;$$

then if follows from (8) or (14) that

$$\text{Degree of polarization} = \frac{\left[n_r^{(1)}(\theta)^2 - n_r^{(2)}(\theta) \right]^2}{\left[n_r^{(1)}(\theta)^2 + n_r^{(2)}(\theta) \right]^2} \; .$$

G. Bekefi

References

1. S. M. Rytov, Theory of Electric Fluctuations and Thermal Radiation, P. N. Lebedev Physical Institute, USSR Academy of Sciences (Academy of Sciences Press, Moscow, 1953). (English Translation, Report AFCRC-TR-59-162, Air Force Cambridge Research Center, Bedford, Mass.).

2. F. V. Bunkin, Soviet Phys.–JETP 5, 665 (1957).

3. R. P. Mercier and H. M. Wills, Physics Laboratory, University of Bristol (unpublished report, 1962).

4. V. I. Pakhomov, V. F. Aleksin, and K. N. Stepanov, Soviet Phys.–Tech. Phys. 6 856 (1962).

5. W. P. Allis, S. J. Buchsbaum, and A. Bers, Waves in Anisotropic Plasmas (The M. I. T. Press, Cambridge, Mass., 1963).

F. APPLICATION OF PARALLEL PLATE GEOMETRY TO THE REFLECTION OF GUIDED WAVES FROM A MAGNETIZED PLASMA

Experimental studies of the reflection of guided waves from a longitudinally magnetized plasma column were presented in Quarterly Progress Report No. 76, pp. 81-86. The power reflected during the afterglow of a pulsed argon discharge tube inserted in a section of S-band waveguide was studied as a function of magnetic field, electron density, and incident frequency. In conformity with the theory of propagation of electromagnetic waves in cold plasma, the condition for a maximum or minimum in reflection amplitude depended only on the dimensionless parameters $\alpha^2 = \dfrac{\omega_p^2}{\omega^2}$ and $\beta^2 = \dfrac{\omega_c^2}{\omega^2}$. Furthermore, in the α^2, β^2-plane, the locus of a given maximum (or minimum) formed a fairly straight line starting at the point $\alpha^2 = 0$, $\beta^2 = 1$ and extending in some cases as far as $\alpha^2 = 1$.

It was suggested at the time that reflection maxima could arise from constructive interference of waves off the two faces of the plasma region so that the loci plotted in the α^2, β^2-plane would correspond to curves of $k_z(\alpha^2, \beta^2) = $ constant, where k_z is the

wave number along the waveguide. Further experimental studies have shown, however, that the reflection characteristics at low densities and magnetic fields are determined primarily by a cavity mode that exists in the empty waveguide. The cylindrical metal sleeves or "chokes" which surround the plasma column outside the guide and prevent radiation losses, provide a chamber wide enough to support a trapped standing wave. The electric field of this mode is perpendicular to that of the normal guided wave and so is cut off in the region beyond the chokes. Resonances of the empty "cavity" have been found at frequencies of 2534, 3067, and 3480 Mc. At other frequencies an incident propagating wave is undisturbed but, just on either side of resonance, the impedance mismatch is considerable. This explains the double peaks in reflection observed near these frequencies at low electron densities.

The problem of understanding the reflection characteristics at high densities and at magnetic fields close to cyclotron resonance still remains. In this region, the plasma column itself could have an appreeiable index of refraction and become the principal reflecting agent. Indeed, the observed effect is a series of equally spaced maxima and minima strongly suggestive of an interference pattern. Partial agreement may be obtained by simply using the relation $k_z(a^2, \beta^2) = $ constant, where $k_z^2 = k_r^2 = \dfrac{\omega^2}{c^2}\left[1 - \dfrac{a^2}{1 - \beta}\right]$; that is, by considering a right circularly polarized wave propagating along the magnetic field. Unfortunately, pure circular waves cannot exist in a waveguide with perfectly conducting walls and a less more realistic model is required. The simplest geometry that includes guide effects is that of two parallel conducting plates infinite in extent and uniformly filled with homogeneous plasma. Consider the applied magnetic field, B_o, to be in the z-direction, which is parallel to the plates. The field equations governing this case have been presented by A. Bers.[1] Briefly, assuming the longitudinal and time dependence of the fields to be $e^{i(\omega t - k_z z)}$, a fourth-order scalar wave equation involving only transverse coordinates is obtained for the z-component of either $\vec{E}$ or $\vec{H}$. The particular advantage of the parallel plate geometry is that the wave functions are merely sums of sines and cosines and one can turn his attention immediately to the dispersion relation.

The solutions split into two separate clasese, symmetric and antisymmetric with respect to a plane midway between the plates. For the symmetric case, application of the condition that the tangential component of $\vec{E}$ vanish at the walls yields the dispersion relation

$$f_1 p_2 \cos(p_1 d) \sin(p_2 d) - f_2 p_1 \cos(p_2 d) \sin(p_1 d) = 0 \tag{1}$$

Here 2d is the plate separation, p_1 and p_2 are transverse wave numbers given by the two roots of

$$p^4 + \left[k_o^2 \left(K_{\parallel} + \frac{K_r K_{\ell}}{K_{\perp}} \right) - k_z^2 \left(\frac{K_{\parallel}}{K_{\perp}} + 1 \right) \right] p^2 + \frac{K_{\parallel}}{K_{\perp}} \left(k_o^2 K_r - k_z^2 \right) \left(k_o^2 K_{\ell} - k_z^2 \right) = 0, \qquad (2)$$

and the f_i are defined by

$$f_i = \frac{K_{\perp}}{k_z K_{\times}} \left[\frac{\left(k_o^2 K_{\perp} - k_z^2 \right) p_i^2}{\left(k_o^2 K_r - k_z^2 \right) \left(k_o^2 K_{\ell} - k_z^2 \right)} - 1 \right].$$

In the above we have used $k_o^2 = \dfrac{\omega^2}{c^2}$ as the frequency variable. The capital K's are various dielectric coefficients related to α and β by

$$K_r = 1 - \frac{\alpha^2}{1 \mp \beta}, \quad K_{\perp} = 1 - \frac{\alpha^2}{1 - \beta^2}, \quad K_{\parallel} = 1 - \alpha^2 \text{ and } K_{\times} = \frac{i\alpha^2 \beta}{1 - \beta^2}.$$

The above equations present a rather involved relation between the four variables α, β, k_o, and k_z. In principle, numerical results can be obtained by fixing the values of three of these parameters and varying the fourth until Eq. 1 is satisfied. Usually an iterative scheme of successive approximations based on the slope of the function in question is employed, but for the method to be effective, a reasonable initial guess as to the approximate location of a desired root must be made. This is especially true in the present problem with its discrete spectrum of transverse eigenfunctions providing an infinitude of roots that must be distinguished. In order to simplify the dispersion relation further and understand the various types of propagating modes, it is convenient to consider the limiting case in which $\alpha^2 \to 0$, $\beta^2 \to 1$ but the ratio $\dfrac{\alpha^2}{\beta^2 - 1} = \eta$ is arbitrary. The use of this limit is prompted by the fact that the observed loci of reflection maxima correspond roughly to the condition $\eta = $ constant. The dielectric coefficients then take on the limiting values

$$K_r = 1 + 2\eta, \quad K_{\ell} = K_{\parallel} = 1, \quad K_{\perp} = 1 + \eta \text{ and } K_{\times} = -i\eta.$$

Equation (2) for the p_i becomes

$$p^4 - \left[k_o^2 \left(1 + \frac{1+2\eta}{1+\eta} \right) - k_z^2 \left(\frac{1}{1+\eta} + 1 \right) \right] p^2 + \frac{1}{1+\eta} \left[k_o^2 (1+\eta) - k_z^2 \right] \left[k_o^2 - k_z^2 \right] = 0,$$

which factors directly giving

$$p_1^2 = k_o^2 - k_z^2 \quad \text{and} \quad p_2^2 = \frac{(1+2\eta) k_o^2 - k_z^2}{1 + \eta}. \qquad (3)$$

Finally, it turns out that

$$f_1 p_z = \frac{-k_o^2}{k_z i_z} \quad \text{and} \quad f_2 p_1 = \frac{-k_z}{p_1}$$

so that Eq. (1) simplifies to

$$p_1 \cos (p_1 d) \sin (p_z d) + \frac{k_z^2}{k_o^2} p_z \cos (p_z d) \sin (p_1 d) = 0 \qquad (4)$$

with p_1 and p_2 given in (3). A similar relation with the sines and cosines interchanged applies to the antisymmetric case. In the regions where either p_1^2 or p_2^2 is negative, hyperbolic trigonometric functions are used. However, if both p_1 and p_2 are pure imaginary, Eq. 4 becomes

$$|p_1| \cosh (|p_1| d) \sinh (|p_z| d) + \frac{k_z^2}{k_o^2} |p_2| \cosh (|p_2| d) \sinh (|p_1| d) = 0,$$

which has no solution for real k_o and k_z.

Figure X-19 shows a diagram of the η, $\left(\frac{k_z}{k_o}\right)^2$-plane indicating the areas where propagation is possible. The line $\eta = 0$ corresponds to the empty waveguide and the line $\eta = -1$ to the plasma resonance condition $\frac{K_r K_\ell}{K_\perp} = \frac{1 + 2\eta}{1 + \eta} = \infty$. The left boundary provides the cutoff limit $k_z = 0$, and the line $\left(\frac{k_z}{k_o}\right)^2 = 1$ represents waves traveling with a phase velocity in the z-direction equal to the speed of light. All modes for $-1 < \eta < 0$ approach this limit as $k_o \to \infty$. Modes for $\eta > 0$ are seen to approach the limit $k_z^2 \to (1+2\eta)k_o^2$ but for $\eta < -1$, the ratio k_z/k_o can become arbitrarily large.

The cutoff limit $k_z/k_o \to 0$ is easily taken. For symmetric modes Eq. 4 becomes $p_1 \cos (p_1 d) \sin (p_2 d) = 0$ and yields two solutions

$$\cos (p_1 d) = 0, \quad \text{or} \quad k_o d = (2n-1) \frac{\pi}{z}$$

and
$$n = 1, 2, 3, \ldots$$

$$\sin (p_2 d) = 0 \quad \text{or} \quad k_o d = \sqrt{\frac{1+\eta}{1+2\eta}} n\pi$$

The first solution corresponds to the TM or "ordinary" cutoff whereas the second gives the TE or "extraordinary" cutoff condition.

With the aid of a digital computer one can now proceed as follows: For a given value of η (starting at $k_z = 0$ and k_o = cutoff limit of desired mode) k_z can be increased by
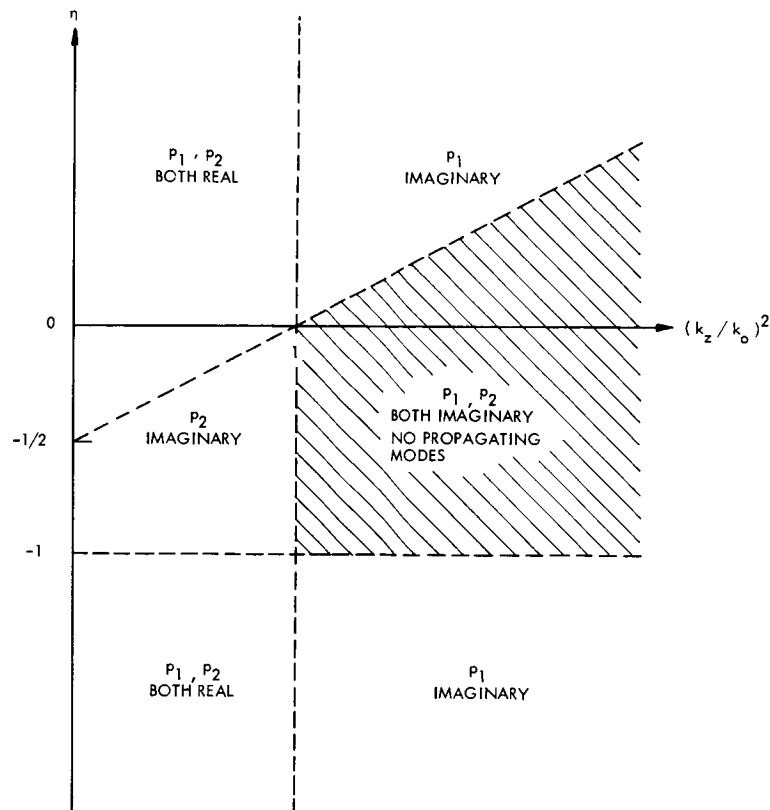

Fig. X-19.   Properties of p_1 and p_2 in the

η, $(k_z/k_d)^2$-plane.

increments and the iterative process employed to trace an entire curve relating k_o to k_z. Such a curve is not a proper dispersion relation because a and β are not varied along with k_o, but a collection of such curves provides the necessary starting values for proceeding away from the limit $a^2 = 0$, $\beta^2 = 1$ out into the rest of the a^2, β^2-plane.

Typical results of this method are shown in Figs. X-20 and X-21 where curves corresponding to various values of k_z are presented for a fixed value of k_o.   These should be compared with the experimentally observed loci of reflection maxima given in the previous report.[2] Figure X-20 is based on the lowest antisymmetric mode, which has the cutoff properties of a TE wave.   No relation between the traces in Fig. X-20 and the experimental ones is evident.   Figure X-21, which illustrates the lowest symmetric mode, is far more encouraging.   At small values of a^2, these curves exhibit all the properties of the observed loci including the very crucial one that no curves lie in the
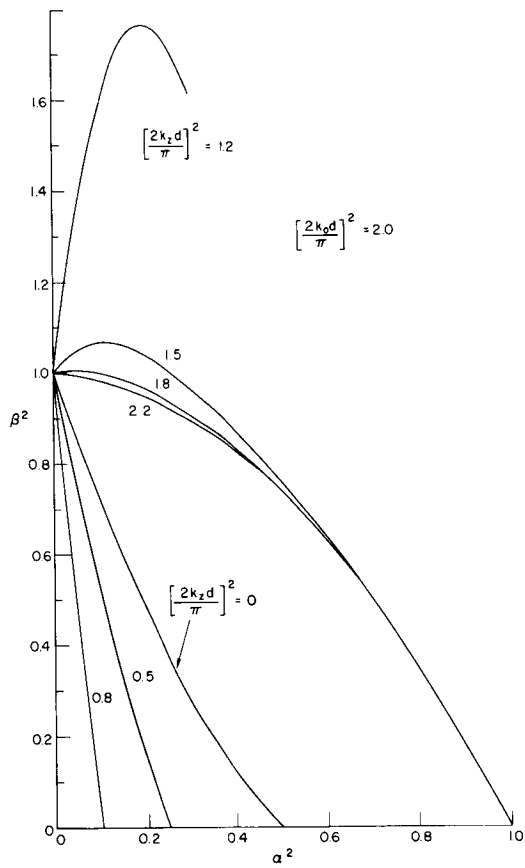
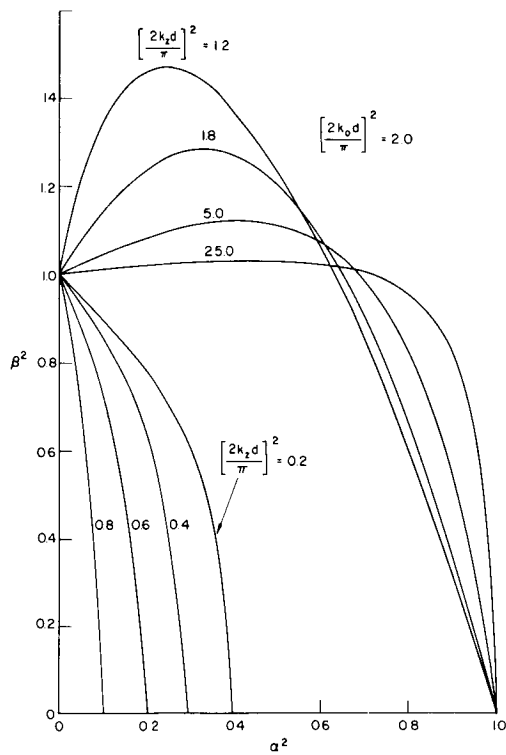Fig. X-20. Lines of constant k_0 and k_z for the antisymmetric mode.

Fig. X-21. Lines of constant k_0 and k_z for the symmetric mode.

region bounded by $\alpha^2+\beta^2 = 1$ and $\beta^2 = 1$. The only shortcoming is the bending of the curves as α^2 becomes significant; no such curvature was found in the experiment. In this connection it should be pointed out that the reported electron density was that of the plasma column itself and hence was greater than an average taken over the waveguide cross section. Furthermore, plots similar to those shown in Fig. X-20, taken at various values of k_o, show that as k_o increases the curves are straight over a larger range of α^2 before bending.

<div align="right">B. L. Wright</div>

References

1. A. Bers in W. P. Allis, S. J. Buchsbaum and A. Bers, Waves in Anisotropic Plasmas (The M. I. T. Press, Cambridge, Mass., 1963), Section 10. 3. 1.

2. B. L. Wright, Quarterly Progress Report No. 76, Research Laboratory of Electronics, M. I. T., January 15, 1965, Fig. XIV-3, p. 83.

Prof. L. D. Smullin
Prof. H. A. Haus
Prof. A. Bers
Prof. W. D. Getty
Prof. D. J. Rose
Prof. T. H. Dupree
Prof. L. M. Lidsky
Prof. E. P. Gyftopoulos
Dr. T. Musha
R. R. Bartsch
T. S. Brown
J. F. Clarke
J. A. Davis
F. E. Dunn

P. H. Edmonds
S. A. Evans
T. J. Fessenden
R. W. Flynn
F. Gardiol
E. T. Gerry
J. N. Hamawi
C-F. G. Hsi
G. I. Kachen, Jr.
C. A. Kapentanakos
B. R. Kusse
S. H. Kyong
M. A. Lieberman

M. Lind
M. D. Lubin
J. D. Mills
R. W. Moir
A. A. Offenberger
K. C. Papadopoulos
R. R. Parker
L. M. Petrie, Jr.
C. S. Ribbeck
H. M. Schneider
E. Thompson
C. E. Wagner
R. N. Wallace
J. C. Woo

A. ION-PLASMA OSCILLATIONS

The plasma density of the electron-cyclotron resonance discharge (ECRD) has been measured with an 8-mm Fabry-Perot microwave interferometer. The ECRD experiment has been described in Quarterly Progress Report No. 73 (pages 81-85) and Quarterly Progress Report No. 75 (pages 120-121). The discharge is driven by a 2.4-Gc, 1.2-kw magnetron, powered by an unfiltered 5-kv power supply. The magnetron current and voltage waveforms are shown in Fig. XI-1.
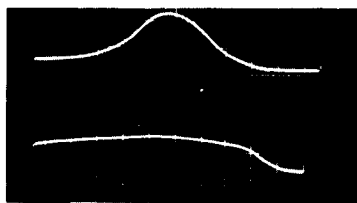
Fig. XI-1. Magnetron current and voltage vs time. Upper trace: magnetron current, 0.1 amp/cm. Lower trace: magnetron voltage, 5000 volts/cm. Time scale, 0.5 msec/cm.

The Fabry-Perot interferometer has been described in Quarterly Progress Report No. 76 (pages 109-111). It consists of a pair of spherical copper mirrors mounted facing each other in the side wall of the discharge tube. The mirrors are $3\frac{1}{2}$ inches in diameter and approximately 6 inches apart. The mirror system is confocal, that is, each mirror is focused at the center of the discharge tube. A 35-Gc, 10-mw klystron provides the power for this system. The microwave energy is fed from the 8-mm waveguide into

---

and out of the resonating system through a small hole in the center of each mirror. A crystal is used to detect the microwave energy passing through the resonating system.

The klystron is modulated by a sawtooth whose period τ_s is short compared with the width of the magnetron current pulse. Typically, τ_s is approximately 3.5 μsec. As the klystron is swept in frequency, a fringe or "pip" is detected and displayed on an oscilloscope when the klystron frequency coincides with a resonant frequency of the mirror system. As the plasma builds up inside the discharge tube, the position of this fringe shifts. By measuring this shift, the electron density of the plasma may be determined.

The interferometer phase shift as a function of time has been measured for various values of the pressure and magnetic field in a hydrogen discharge. The electron density was then calculated, under the assumption that the plasma uniformly filled the discharge tube. The actual density profile in the discharge tube has not yet been determined. In Fig. XI-2, a typical plot of the electron density against time is shown. Each point on the plot represents a separate magnetron current pulse. This pulse is shown in Fig. XI-2 as a dotted line. Even though the shape and height of the magnetron current do not vary from pulse to pulse, the pulse-to-pulse variation of the plasma density at any time within a magnetron current pulse is considerable. This indicates that the plasma density is
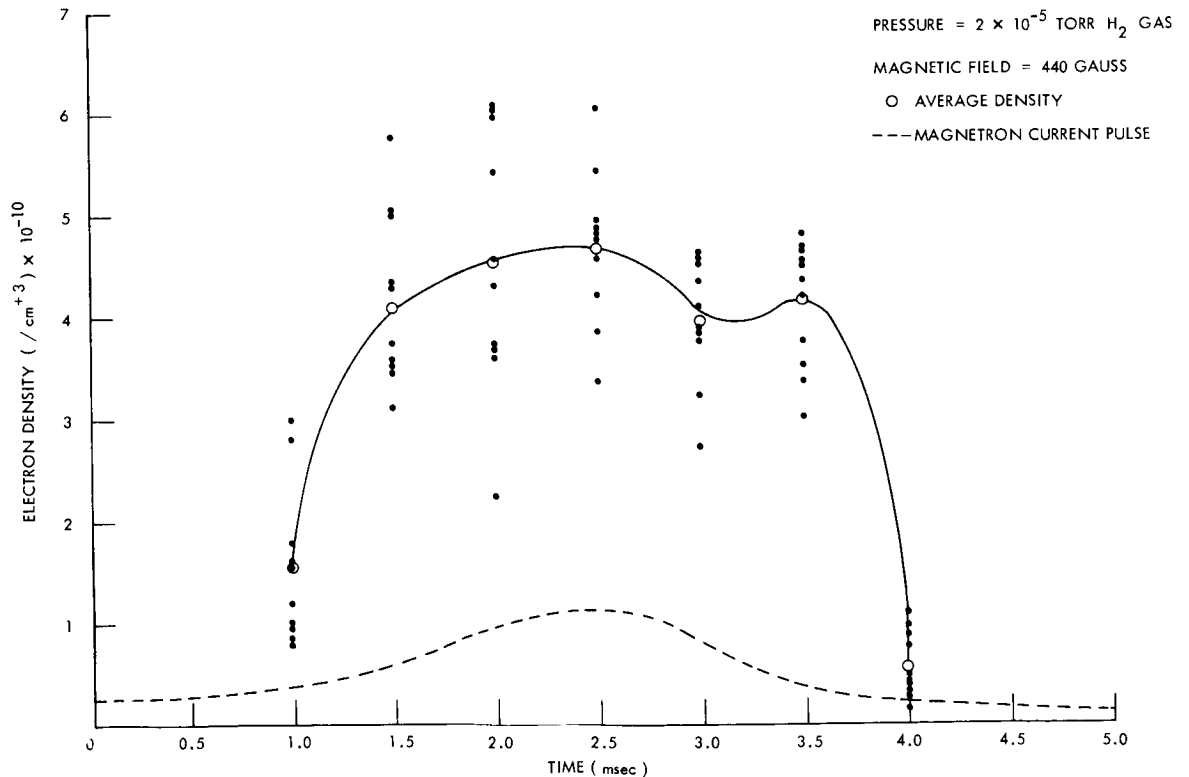

Fig. XI-2.   Electron density vs time in ECRD.

fluctuating considerably (by 50 per cent or more) in the ECRD, and on a time scale that is long compared with τ_s, the interferometer sweep time. The fluctuation frequency is thus bounded above by $\frac{1}{\tau_s} \approx 30$ kcps and below (presumably) by 60 cps.
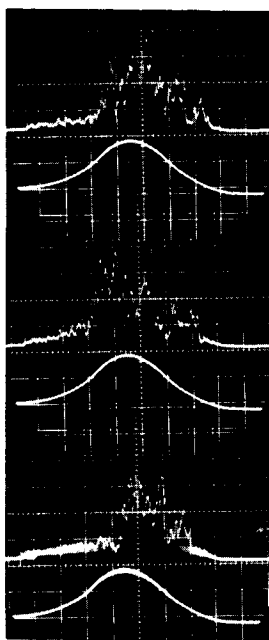
Fig. XI-3.   Retarding potential probe plate current vs time (p = 4×10^{-6} torr Hg, $B_o = 400$ gauss). Upper trace: retarding potential probe plate current. Lower trace: magnetron current pulse. Time scale 0.5 msec/cm.

In order to examine the fluctuating plasma in more detail, the current reaching the plate of a retarding potential probe was measured as a function of time. The retarding potential probe is mounted on the axis of the discharge tube just back of the magnetic mirror. The probe consists of a capped copper tube, $\frac{3}{4}$ inch in diameter and approximately 2 inches long. A hole of 25-mil diameter has been drilled in the cap to sample the plasma. Behind the cap are arranged three grids and a plate. In this experiment, the grids were all grounded and the plate is held at +180 volts.

Figure XI-3 shows three separate magnetron current pulses and the corresponding plate current of the retarding potential probe. For these three pulses, the magnetic field in the center of the discharge tube was 440 gauss and the pressure was 4×10^{-6} torr. The fluctuations in the plate current provide a strong indication that the plasma density itself is fluctuating.

The fluctuating frequency can be estimated from Fig. XI-3 to be approximately 6000 cps.

In Fig. XI-4 the average density of the plasma at peak magnetron current (2.5 msec) has been plotted against the magnetic field in the center of the discharge tube, with pressure as a parameter. Note that the magnetic-mirror ratio for this experiment is approximately 4. The lower dotted line is the electron
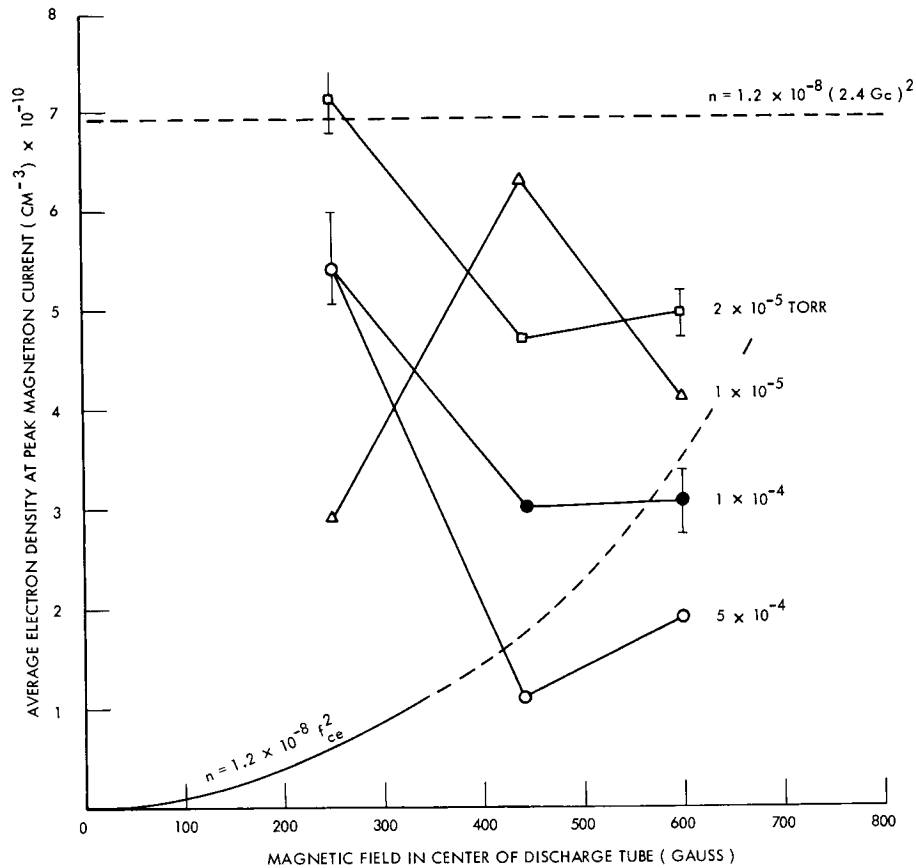
Fig. XI-4. Average electron density vs magnetic field.

density found by assuming that the electron cyclotron and electron plasma frequencies are equal at the center of the discharge tube. The upper dotted line gives the density corresponding to a plasma frequency of 2.4 Gc, the driving frequency of the magnetron. It appears that the average electron density is limited by the magnetron frequency (upper dotted line). Several more points must be taken, however, to complete Fig. XI-4.

M. A. Lieberman

B. DISPERSION DIAGRAMS FOR HOT-ELECTRON PLASMAS

1. Computer Program for Solving Transcendental Dispersion Equations

A computer program has been written to find the zeros of a transcendental dispersion function $\underline{D}(\underline{\omega}, k, \ldots)$ in the complex $\underline{\omega}$ plane. This program must be used in conjunction with the Project MAC time-sharing system.

The user must provide a subroutine that computes the value of $\underline{D}$, given the complex frequency $\underline{\omega}$, wave number k and any other parameters that the user may desire. The subroutine may be written by using MAD, FORTRAN or FAP programming languages.

In operation, the program continuously steps k by an increment Δk, and at each step finds a zero, $\underline{\omega}_n$, of the dispersion function $\underline{D}$

$$\underline{D}(\underline{\omega}_n, k+n\Delta k, \ldots) = 0.$$

The zero, $\underline{\omega}_n$, is found by constructing a grid of values in the complex ω plane around an initial guess ω_{gn} as follows:

$$\begin{cases} \underline{\omega}_{go} = \text{provided by the user} \\ \underline{\omega}_{g1} = \underline{\omega}_o \\ \omega_{gn} = \underline{\omega}_{n-1} + \Delta k(\underline{\omega}_{n-1} - \underline{\omega}_{n-2}) \qquad n \geq 2. \end{cases}$$

The program evaluates the real function $F = |D(\underline{\omega}, k+n\Delta k, \ldots)|^2$ for every point on the grid. If a minimum of F is found at ω_n, the grid is refined several times until $\underline{\omega}_n$ is given to three significant figures; then $\underline{D}$ is checked to verify that its real and imaginary parts change sign in the neighborhood of $\underline{\omega}_n$. If such is the case, the zero is printed and k is stepped. Thus the zeros of D in the complex $\underline{\omega}$ plane are computed as a function of the wave number k.

Man-machine interaction plays an important role in the use of this program. It would be practically impossible to utilize this program in an efficient manner without the aid of the time-sharing system. The grid size and spacing in the complex $\underline{\omega}$ plane, the wave number k and its increment Δk, and all other parameters are specified by the user and may be altered at will. If at any step a zero of $\underline{D}$ is not found, the program requests the user to change the grid size, spacing, and location in the complex $\underline{\omega}$ plane. As an aid to the user, the program will print the values of $\underline{D}$ and F at the grid points if desired.

2. Longitudinal Waves in the Absence of a Magnetic Field

The dispersion equation for longitudinal waves propagating in a hot-electron Maxwellian plasma[1] is

$$1 + \frac{1}{k^2 \lambda_D^2} \left[1 + \frac{1}{\sqrt{2}\, k\lambda_D} \frac{\omega}{\omega_{pe}} Z\left(\frac{1}{\sqrt{2}\, k\lambda_D} \frac{\omega}{\omega_{pe}} \right) \right] = 0,$$

where ω_{pe} is the electron plasma frequency, $\lambda_D = \dfrac{V_T}{\omega_{pe}}$ is the Debye wavelength, and

$$Z(\xi) = \pi^{-1/2} \int_{-\infty}^{\infty} dx \, \frac{e^{-x^2}}{x - \xi} \qquad \mathrm{Im}\ \xi > 0$$

is the plasma dispersion function, tabulated by Fried and Conte.[2]

The solution of this dispersion equation is shown in Fig. XI-5. The familiar Landau damping result is obtained for $k\lambda_D \ll 1$, while for $k\lambda_D$ much in excess of 1, the wave is heavily damped.
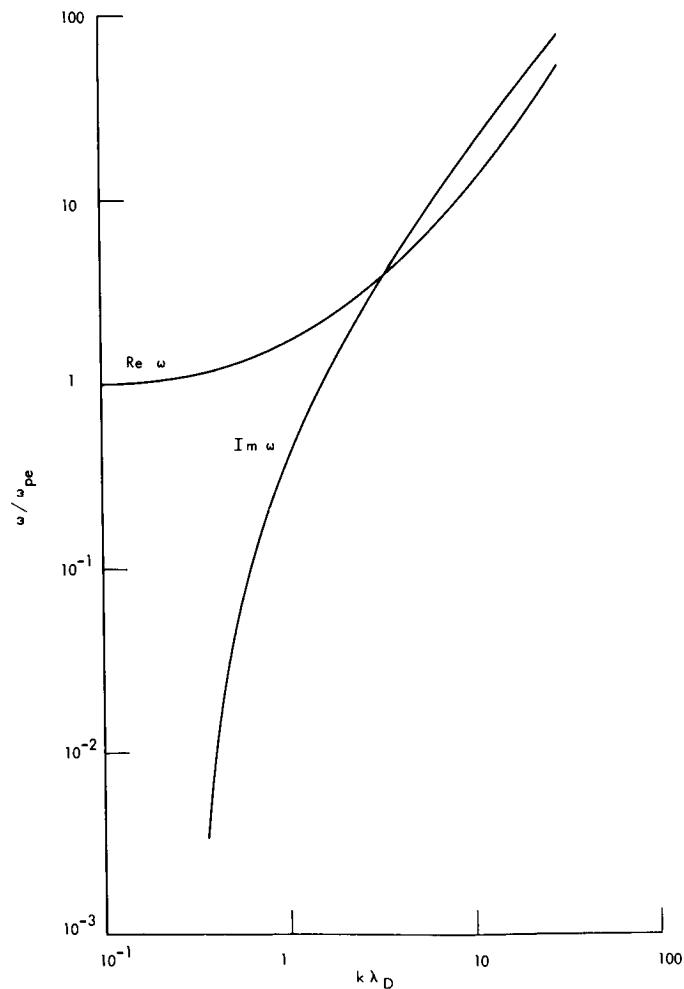
Fig. XI-5. Plasma dispersion equation for longitudinal waves.

3. Quasi-static Waves in a Hot-Electron Plasma Waveguide

Consider a hot-electron plasma waveguide whose axis is parallel to the static magnetic field. Under the assumption that as a boundary condition the tangential electric
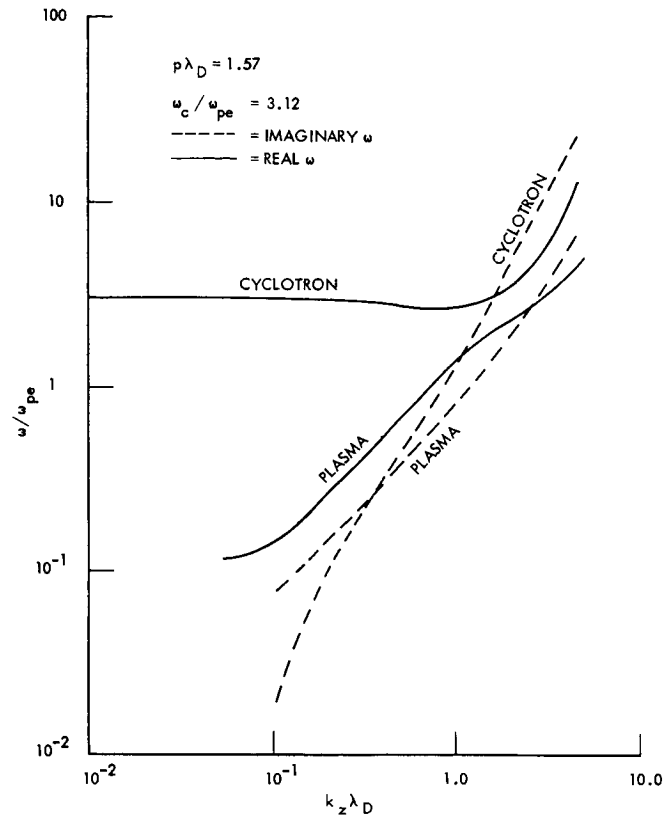
Fig. XI-6. Quasi-static dispersion equation in a hot-electron plasma waveguide.

field vanishes at the waveguide walls, the quasi-static dispersion equation for this system[1] is

$$1 + \frac{1}{\left(k_z^2 + p^2\right)\lambda_D^2}\left[1 + \xi_0 \sum_{n=-\infty}^{\infty} Z(\xi_n)\, I_n(\lambda)\, e^{-\lambda}\right] = 0,$$

where $\xi_n = \left(\dfrac{1}{\sqrt{2}\, k_z \lambda_D}\right)\left(\dfrac{\omega}{\omega_{pe}} - n\dfrac{\omega_c}{\omega_{pe}}\right)$, $\lambda = p\lambda_D \Big/ \left(\dfrac{\omega_c}{\omega_{pe}}\right)$, ω_c is the electron cyclotron frequency, k_z is the longitudinal wave number, p is the transverse wave number, and $I_n(x)$ is the Bessel function of imaginary argument.

$$I_n(x) = i^{-n}\, J_n(ix).$$

This dispersion equation has an infinite number of solutions, each one corresponding roughly to a cyclotron frequency harmonic. For these solutions, $\omega \to n\omega_c$ as $k_z \to 0$. In addition to these waves, there is a single plasma wave, for which $\omega \to 0$ as $k_z \to 0$.

Through the use of the computer program described above the two solutions for the plasma wave and the first-harmonic cyclotron wave have been obtained. Figure XI-6 shows the solutions for the parameters $p\lambda_D = 1.57$, $\omega_c/\omega_{pe} = 3.12$. There is a crossover point at $k_z\lambda_D = 0.35$. Below this wave number, the cyclotron wave dominates the behavior of the waveguide system, since it has the smaller loss. Both waves are damped out for $k_z\lambda_D$ much greater than 2.5.

It should be noted that quasi statics is invalid in the neighborhood of the electron cyclotron frequency and its harmonics. Also, the question of boundary conditions arises. In a plasma with transverse, as well as longitudinal, temperature, plasma particles are constantly hitting the walls. It is not clear whether the boundary condition considered here, that the tangential electric field be zero on the waveguide wall, even approximates the actual plasma boundary condition.

M. A. Lieberman

References

1. T. Stix, The Theory of Plasma Waves (McGraw-Hill Publishing Company, New York, 1962), p. 103.

2. B. Fried and S. Conte, The Plasma Dispersion Function (Academic Press, New York, 1961).

C. INSTABILITIES IN ELECTRON STREAMS IN CROSSED ELECTRIC AND MAGNETIC FIELDS

We have been studying the properties of a hot, weak-density, unneutralized electron beam, whose unperturbed motion consists of a drift perpendicular to crossed static electric and magnetic fields.

In the unperturbed state, we assume that the average velocity of the electrons is given by $\overline{v}_{ave} = v_{0x}\overline{i}_x$, where $v_{0x} = E_0/B_0$. (See Fig. XI-7.) Here $E_0\overline{i}_y$ and $B_0\overline{i}_z$ are the applied, static electric and magnetic fields. We neglect the DC space charge and current, and the zero-order fields produced by them. The Boltzmann equation for this unperturbed state is satisfied by the distribution function

$$f_0 = A \exp\left[-\frac{1}{2}m\frac{(v_x-v_{0x})^2}{k_BT}\right]\exp\left[-\frac{1}{2}\frac{mv_y^2}{k_BT}\right]g(v_z), \tag{1}$$

where $v_{0x} = E_0/B_0$, A is a normalization constant, v_x, v_y, and v_z are the x, y, and
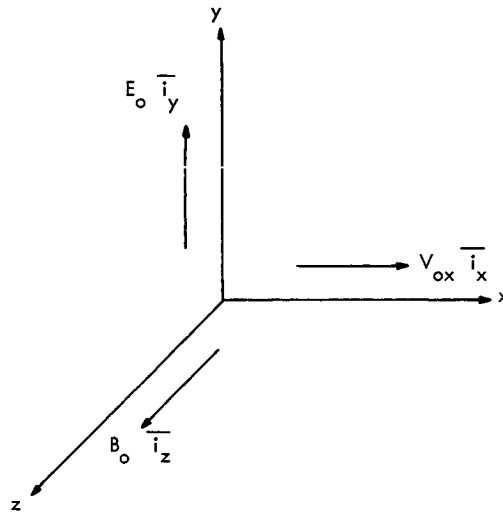
Fig. XI-7.   Coordinate system for electron stream in
crossed electric and magnetic fields.

z components of the particle velocity, m is the electron mass, T is the temperature,
and k_B is Boltzmann's constant.   The function $g(v_z)$ is left arbitrary at this point.   Note
that f_0 is spatially homogeneous and independent of time; it satisfies

$$\frac{q}{m}(E_0\bar{i}_y + \vec{v} \times B_0 \bar{i}_z) \cdot \nabla_v f_0 = 0, \tag{2}$$

where q is the charge of a single electron.

If this unperturbed motion is given a perturbation so that the distribution function

$f = f_0 + \mathrm{Re}\ [f_1\ e^{j(\omega t - \bar{k} \cdot \bar{r})}]$, the linearized Boltzmann equation becomes

$$j(\omega - \bar{k} \cdot \bar{v})\ f_1 + \frac{q}{m}\ E_0 \frac{\partial f_1}{\partial v_y} + \frac{q}{m}\ (\vec{v} \times B_0 \bar{i}_z) \cdot \nabla_v f_1 = -\frac{q}{m}\ \bar{E}_1 \cdot \nabla_v f_0. \tag{3}$$

Here $\bar{E}_1$ is the perturbed electric field.   We shall make the quasi-static assumption,
namely that the perturbed magnetic field $\bar{B}_1 = 0$, and hence $\bar{E}_1 = -\nabla\Phi$.   Equation 3 is
difficult to solve, however, owing to the term involving E_0.

If a Galilean transformation is made to a coordinate system translating with the
average velocity $v_{0x}\bar{i}_x$ of the electrons, the static electric field vanishes and the prob-
lem is considerably simplified.   This transformation may be described by the relations

$$\vec{r} = \vec{r}' + v_{0x}\bar{i}_x t \tag{4}$$

$$\vec{v} = \vec{v}' + v_{0x}\bar{i}_x \tag{5}$$

$$t = t' \tag{6}$$

$$\overline{E} = \overline{E}' - v_{0x}\overline{i}_x \times \overline{B}' \tag{7}$$

$$\overline{B} = \overline{B}'. \tag{8}$$

The symbol prime denotes quantities measured in the moving system, while unprimed quantities are measured in the laboratory system. In Eqs. 4-8, we have assumed $\dfrac{v_{0x}}{c} \ll 1$, where c is the velocity of light. In the moving coordinate system, the Boltzmann equation becomes

$$\frac{\partial f(\overline{r}', \overline{v}', t')}{\partial t'} + \overline{v}' \cdot \nabla_{r'} f(\overline{r}', \overline{v}', t') + \frac{q}{m}\left[\overline{E}' + \overline{v}' \times \overline{B}'\right] \cdot \nabla_{v'} f(\overline{r}', v', t',) = 0. \tag{9}$$

The unperturbed distribution function is

$$f = A\,\exp\left[-\frac{1}{2}\frac{mv'^2_x}{k_B T}\right]\exp\left[-\frac{1}{2}\frac{mv'^2_y}{k_B T}\right]g(v'_z) \tag{10}$$

which satisfies the unperturbed Boltzmann equation

$$\frac{q}{m}(\vec{v}' \times B_0\overline{i}_z) \cdot \nabla_{v'} f_0 = 0. \tag{11}$$

The linearized Boltzmann equation (if we assume an $e^{j(\omega' t' - \overline{k}' \cdot \overline{r}')}$ dependence for f_1) is

$$j(\omega' - \overline{k}' \cdot \overline{v}')\,f_1 + \frac{q}{m}(\vec{v}' \times B_0\overline{i}_z) \cdot \nabla_{v'} f_1 = -\frac{q}{m}\overline{E}'_1 \cdot \nabla_{v'} f_0. \tag{12}$$

The dispersion relation[1] is found to be

$$0 = 1 + \frac{\omega_p^2}{k'^2}\int_0^\infty v'_\perp dv'_\perp \int_0^\infty dv'_z \sum_{m=-\infty}^{m=+\infty} \frac{J_m^2(p)}{(\omega' - k'_\parallel v'_z - m\omega_c)}\left(\frac{m\omega_c}{v'_\perp}\frac{\partial}{\partial v'_\perp} + k'_\parallel \frac{\partial}{\partial v'_z}\right)\frac{2\pi}{n_0}f_0(v'). \tag{13}$$

Here, if e is the magnitude of the electronic charge and n_0 the number density of electrons, ω_p is the plasma frequency $\left(\omega_p^2 = \dfrac{e^2 n_0}{\epsilon_0 m}\right)$, and ω_c is the cyclotron frequency $\left(\omega_c = \dfrac{eB_0}{m}\right)$. J_m is a Bessel function of the first kind, order m; $v'_\perp = \sqrt{v'^2_x + v'^2_y}$; $p = \dfrac{k'_\perp v'_\perp}{\omega_c}$. The wave number $k'_\parallel$ is the component of $\overline{k}'$ along $B_0\overline{i}_z$, while $k'_\perp = \sqrt{k'^2_x + k'^2_y}$.

In order to find the dispersion relation in the laboratory system, one uses the Galilean transformations for frequency and wave number:

$$\omega = \omega' + k'_x v_{0x} \tag{14}$$

Fig. XI-9. Dispersion relation showing instability at $\omega' \approx 0.6\,\omega_c$.

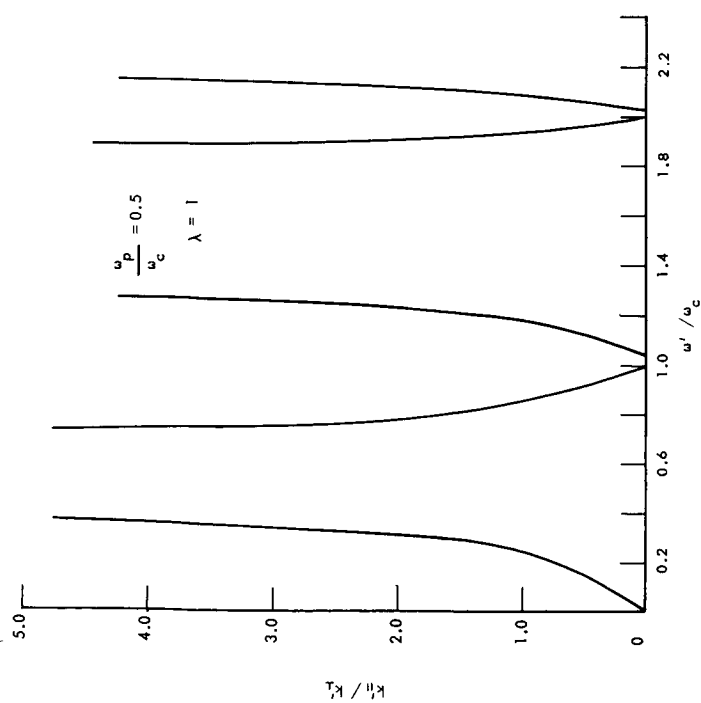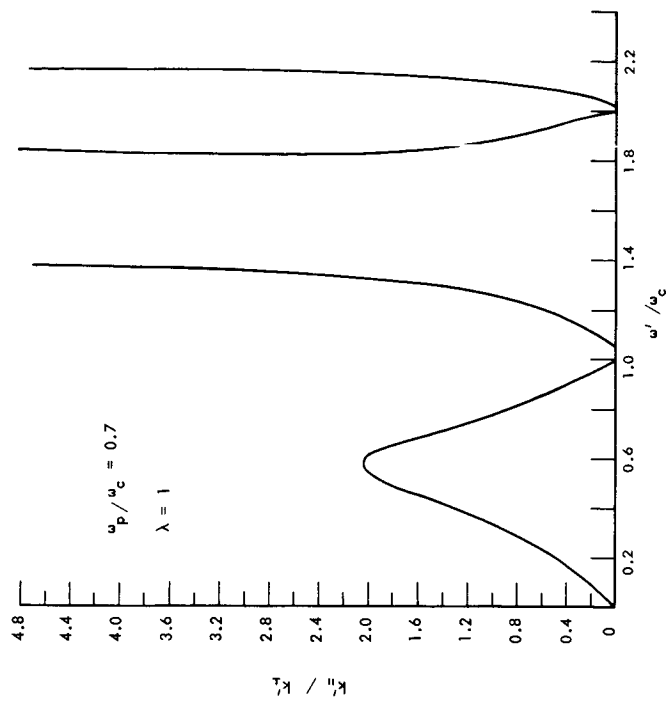

Fig. XI-8. Dispersion relation showing stable solutions.

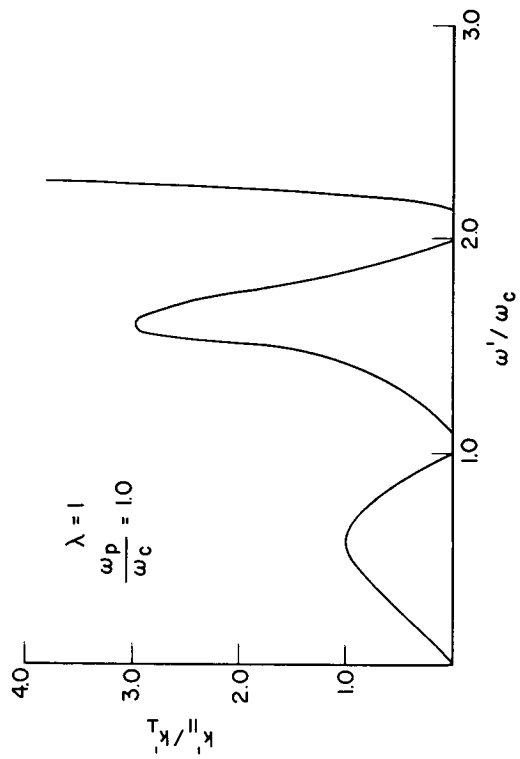Fig. XI-10. Dispersion relation showing instabilities at $\omega' \approx 0.6\,\omega_c$ and $\omega' \approx 1.6\,\omega_c$.

$$\overline{k} = \overline{k}'.$$

(15)

If the unperturbed distribution function for the beam is taken as in Eqs. 1 and 10, with $g(v_z') = \delta(v_z')$, the dispersion relation becomes

$$\frac{k'^2_{\parallel}}{k'^2_{\perp}} = -\frac{1 - \dfrac{\omega_p^2}{\omega_c^2} e^{-\lambda} \displaystyle\sum_{n=-\infty}^{n=+\infty} I_n(\lambda)\, \dfrac{n/\lambda}{\left(\dfrac{\omega'}{\omega_c} - n\right)}}{1 - \dfrac{\omega_p^2}{\omega_c^2} e^{-\lambda} \displaystyle\sum_{n=-\infty}^{n=+\infty} I_n(\lambda)\, \dfrac{1}{\left(\dfrac{\omega'}{\omega_c} - n\right)^2}},$$

(16)

where $I_n(x)$ is a modified Bessel function of nth order and the quantity $\lambda = \dfrac{k'^2_{\perp}}{\omega_c^2}\left(\dfrac{k_B T}{m}\right)$.

In Figs. XI-8, XI-9, and XI-10 we show three plots of $\dfrac{k'_{\parallel}}{k'_{\perp}}$ versus the normalized frequency $\dfrac{\omega'}{\omega_c}$ for different values of $\dfrac{\omega_p}{\omega_c}$ at fixed $\lambda = 1.0$. For sufficiently small values of density $\dfrac{\omega_p}{\omega_c} < 0.7$, all values of $\dfrac{k'_{\parallel}}{k'_{\perp}}$ are stable. This may be seen from Fig. XI-8, where $\dfrac{\omega_p}{\omega_c} = 0.5$. When $\dfrac{\omega_p}{\omega_c} = 0.7$ (see Fig. XI-9), waves with $\dfrac{k'_{\parallel}}{k'_{\perp}} > 2.1$ are unstable, and the real part of the frequency lies in the range $0 < \dfrac{\omega'}{\omega_c} < 1$. As $\dfrac{\omega_p}{\omega_c}$ increases, more instabilities appear. For example, as shown in Fig. XI-10, for $\dfrac{\omega_p}{\omega_c} = 1.0$, waves with $\dfrac{k'_{\parallel}}{k'_{\perp}} > 1.0$ are unstable, with the real part of the frequency $0 < \dfrac{\omega'}{\omega_c} < 2.0$.

Recent work on a different problem whose results are related to these has been reported.[2]

H. M. Schneider, A. Bers

References

1. A. Bers and S. Gruber, Negative-energy plasma waves and instabilities at cyclotron harmonics, Appl. Phys. Letters 6, 27 (January 15, 1965).

2. S. Gruber, M. W. Klein, and P. L. Auer, High Frequency Velocity Space Instabilities, Research Report SRRC-RR-76, Sperry Rand Research Center, Sudbury, Mass., January 1965.

D. INSTABILITIES IN TRANSVERSE WAVES ALONG B_0 FOR BEAM-TYPE DISTRIBUTIONS

The dispersion relation for transverse electromagnetic waves propagating along a DC magnetic field on a plasma beam has been analyzed to uncover the instabilities that may exist.

An infinite system of stationary neutralizing ions penetrated by an electron beam is assumed. The velocity space distribution function of the electrons is taken to be

$$f_0(v_\perp, v_\parallel) = \frac{1}{2\pi v_{0\perp}} \, \delta(v_\perp - v_{0\perp}) \, \delta(v_\parallel - v_{0\parallel}),$$

where $v_\perp$ and $v_\parallel$ are the velocities perpendicular to and along the magnetic field.

A simultaneous solution of the relativistic Boltzmann equation and Maxwell's equations yields the dispersion relation

$$\frac{c^2 k^2}{\omega^2} = 1 - \frac{\omega_p^2}{\omega^2} \left[\frac{\omega - k v_{0\parallel}}{(\omega - k v_{0\parallel} - \omega_b)} + \frac{v_{0\perp}^2 (k^2 - \omega^2/c^2)}{2(\omega - k v_{0\parallel} - \omega_b)^2} \right],$$

where a dependence, $e^{j(\omega t - kx)}$, of the field quantities is assumed, and only right-handed polarized waves are considered.

The criteria of Bers and Briggs were employed to analyze the dispersion relation.[1] Figure XI-11a shows the roots of the dispersion relation as a function of real k in the complex ω-plane. The plots were obtained on a cathode-ray tube by means of a system known as the "Kludge" associated with project MAC.[2] As k goes from $-\infty$ to $+\infty$, the locus in the ω-plane is traced from left to right as indicated in Fig. XI-11a. For large k the locus levels off at a maximum negative ω_i. Following Bers and Briggs, we find that instabilities exist over the range of real frequencies with negative imaginary parts. In order to determine the nature of the instability, the roots of the dispersion equation, for the real frequency range of interest, must be investigated in the complex k-plane.

Figure XI-11b shows the locus of roots in the k-plane as ω is varied from ω_L to ω_R. When two roots from opposite sides of the real k-axis meet, the dispersion equation has a saddle point, and an absolute instability is uncovered. When a root crosses the real axis and stops, a convective instability is indicated.

The nature of the instabilities as the parameters ω_p, $v_{0\perp}$, and $v_{0\parallel}$ are varied has been studied. For nonrelativistic velocities, the absolute instabilities, when they are present, occur at the points indicated by (A) and (B) in Fig. XI-11a. As the plasma frequency is increased the long-wavelength instability moves into the high-frequency region of the short-wavelength instability, as in Fig. XI-12a. For relativistic velocities perpendicular to the magnetic field, the two instabilities merge to form the loops shown in Fig. XI-13a.

Fig. XI-12. (a) Real k-axis contours in the complex
ω-plane.
(b) Complex k-plane showing the abso-
lute instability at A.

Fig. XI-11. (a) Real k-axis contours in the complex
ω-plane.
(b) Complex k-plane showing the abso-
lute instability at A.

Fig. XI-14. Real k-axis contours in the complex ω-plane for $v_{0\parallel} = 0$.

$\dfrac{v_{0\parallel}}{c} = 0$

$\dfrac{\omega_p}{\omega_b} = .422$

$\dfrac{v_{0\perp}}{c} = .316$

$k = \pm\infty$

$k = 0$

$k = 0$

$k = \pm\infty$

$\dfrac{\omega_r}{\omega_b}$

$\dfrac{\omega_i}{\omega_b}$

(a)

$\dfrac{\omega_p}{\omega_b} = 1.5$

$\dfrac{v_{0\perp}}{c} = 0.948$

$\dfrac{v_{0\parallel}}{c} = 0.3$

$\dfrac{\omega_r}{\omega_b}$

$\dfrac{\omega_i}{\omega_b}$

ω_L

B

(b)

$\dfrac{k_r c}{\omega_b}$

$\dfrac{k_i c}{\omega_b}$

Fig. XI-13. (a) Real k-axis contours in complex ω-plane (ultra-relativistic beam).
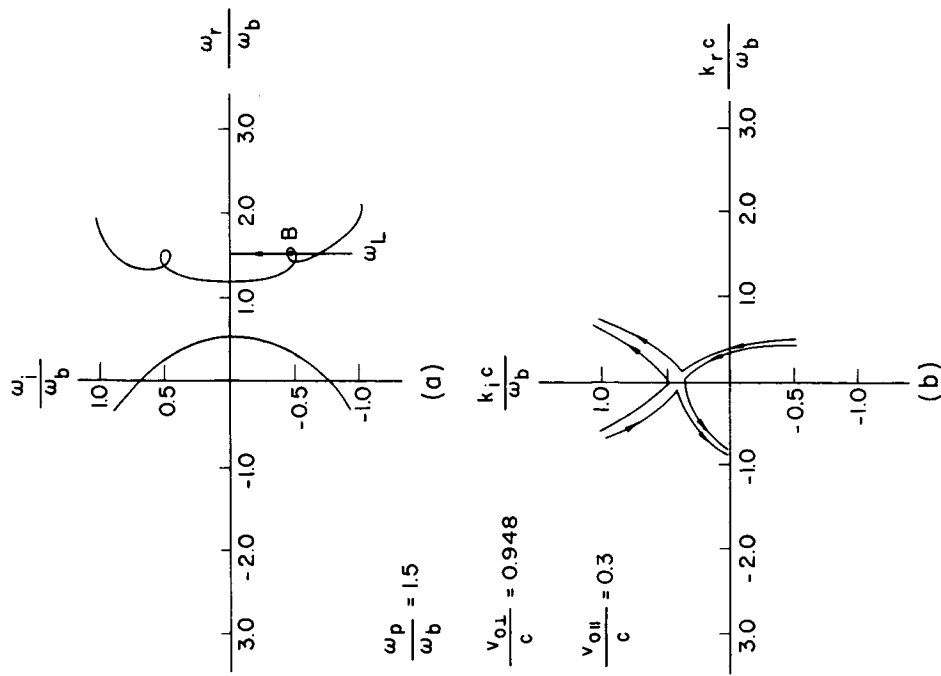(b) Complex k-plane showing the absolute instability at B (ultra-relativistic beam).

For $v_{0\parallel} = 0$, the dispersion relation becomes second-order in ω and the ω-plane plots appear as in Fig. XI-14. For small ω_p the zero and infinite wave number absolute instabilities have the same value of ω_i, and the magnitude increases with ω_p. At $\dfrac{\omega_p}{\omega_0} \sim \dfrac{v_{0\perp}}{c}$ the zero wave number instabilities disappear while the large wave number solutions
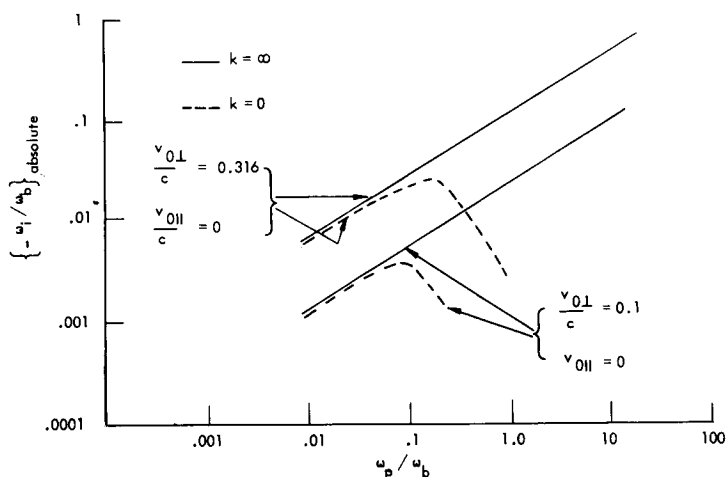
Fig. XI-15. Smooth rate of the absolute instability for $v_{0\parallel} = 0$ as a function of ω_p/ω_b, with $v_{0\perp}/c$ as a parameter.

continue to increase with ω_p. This characteristic is demonstrated in Fig. XI-15. As is also apparent from Fig. XI-15, ω_i increases with $v_{0\perp}$.

The effect of finite temperature on the instabilities discussed here is now under study.

A. Bers, J. K. Hoag, E. A. Robertson

References

1. A. Bers and R. J. Briggs, Criteria for determining absolute instabilities and distinguishing between amplifying and evanescent waves, Quarterly Progress Report No. 71, Research Laboratory of Electronics, M. I. T., October 15, 1963, pp. 122-131.

2. J. Mills and A. Bers, Computer analysis and display of wave instabilities, Quarterly Progress Report No. 73, Research Laboratory of Electronics, M. I. T., April 15, 1964, pp. 85-86.

E. HOT PLASMA WAVEGUIDES AND RESONATORS

1. Model Description

Interactions between electromagnetic fields and plasma in metallic waveguides and cavities are considered in this report. The treatement refers to a hot, collisionless, inhomogeneous electron plasma in a static magnetic field, described by the linearized first two moments of the Vlasov equation and Maxwell's equations. A scalar pressure is assumed and heat flow is neglected.

2. Dielectric-Tensor Operator

The dielectric-tensor operator K_{op} is defined[1] by

$$\overline{\overline{K}}_{op} \cdot \overline{E} = \overline{E} + \frac{\overline{J}}{j\omega\epsilon_0}. \tag{1}$$

For our model of plasma, the following expression is obtained for the dielectric-tensor operator:

$$\overline{\overline{K}}_{op} = \left(\overline{\overline{I}} - \frac{\omega_p^2}{\omega^2} \overline{\overline{A}} \right) + \frac{v_T^2}{\omega^2} n_0 \overline{\overline{A}} \cdot \nabla\left(\frac{1}{n_0} \nabla \right), \tag{2}$$

where

$$\omega_p^2 = \frac{n_0 e^2}{m\epsilon_0} \qquad \text{(electron plasma frequency)} \tag{3}$$

$$v_T^2 = \frac{k_B T}{m} \qquad \text{(electron thermal velocity)} \tag{4}$$

$$\overline{\overline{A}} = \begin{bmatrix} \dfrac{1}{1 - \Omega_b^2} & \vdots & \dfrac{j\Omega_b}{1 - \Omega_b^2} & \vdots & 0 \\ \cdots\cdots\cdots & \vdots & \cdots\cdots\cdots & \vdots & \cdots\cdots \\ \dfrac{-j\Omega_b}{1 - \Omega_b^2} & \vdots & \dfrac{1}{1 - \Omega_b^2} & \vdots & 0 \\ \cdots\cdots\cdots & \vdots & \cdots\cdots\cdots & \vdots & \cdots\cdots \\ 0 & \vdots & 0 & \vdots & 1 \end{bmatrix} \tag{5}$$

Here, $\Omega_b = eB_0/m\omega$, and the static magnetic field B_0 is directed along the z-axis.

3. Energy Conservation Principle

For the assumed lossless plasma, there exists a relation[2] between power flow S and energy W stored in a volume τ:

$$S = - \frac{d}{dt} W, \tag{6}$$

$$S = \oint_\sigma [\bar{E} \times \bar{H} + m v_T^2 n \bar{v}] \cdot \bar{i}_n da \tag{7}$$

$$W = \int_V \frac{1}{2} \left[\mu_0 H^2 + \epsilon_0 E^2 + m n_0 v^2 + \frac{m v_T^2}{n_0} n^2 \right] d\tau. \tag{8}$$

Here, the surface σ encloses the volume τ. Note that in these equations n_0 may be a function of position.

4. Boundary Conditions

Applying the uniqueness theorem to our model of plasma, we find that the solution will be unique if the following boundary conditions are specified:

$\bar{i}_n \times \bar{E} = 0$	on part of the cavity or waveguide wall (perfect electric conductor)	(9)
$\bar{i}_n \times \bar{H} = 0$	on the remainder of the cavity or waveguide wall (perfect magnetic conductor)	(10)
$\bar{i}_n \cdot \bar{v} = 0$	on part of the plasma interface	(11)
$n = 0$	on the remainder of the plasma interface.	(12)

Equations 9-12 may be derived by a technique similar to that used by Bobroff and Haus[3] for electron beams. When the boundary conditions (9-12) are satisfied, the dielectric-tensor operator $\overline{\overline{K}}_{op}$ is a Hermitian operator for the cavity or waveguide considered.

For a cavity of volume τ,

$$\int_\tau \bar{E}^* \cdot \overline{\overline{K}}_{op} \cdot \bar{E} d\tau = \int_\tau \bar{E} \cdot \overline{\overline{K}}_{op}^* \cdot \bar{E}^* d\tau. \tag{13}$$

For a waveguide of cross section A,

$$\int_A \bar{E}^* \cdot \overline{\overline{K}}_{op} \cdot \bar{E} da = \int_A \bar{E} \cdot \overline{\overline{K}}_{op}^* \cdot \bar{E}^* da. \tag{14}$$

When the conditions (9-12) are satisfied, the field solutions will be unique, except at resonances, and for systems capable of sustaining isolated plasma waves (electrostatic waves).

5. Approximate Techniques

The introduction of waveguide and cavity boundaries increases the complexity of the field analysis problem, to the point where it will be difficult or impossible to obtain

exact solutions. It is usually convenient, therefore, to use approximate methods.

a. Perturbation Theory

We compare the hot-plasma system, characterized by the dielectric-tensor operator $\overline{\overline{K}}_{op}$, with a closely related system, characterized by the dielectric tensor $\overline{\overline{K}}$ (for the cold-plasma system) and whose field variables have the subscript 0. By manipulation of Maxwell's equations for the two systems, and by using Green's theorems and the boundary conditions, we obtain relations[4] for the characteristic quantities of the system considered:

(a) Resonant frequency of a cavity

$$\omega - \omega_0 = \frac{-\omega\epsilon_0 \int_\tau \left[\overline{E}_0^* \cdot \overline{\overline{K}}_{op} \cdot \overline{E} - \overline{E} \cdot \overline{\overline{K}}^* \cdot \overline{E}_0^*\right] d\tau}{\int_\tau \left[\mu_0 \overline{H}_0^* \cdot \overline{H} + \epsilon_0 \overline{E} \cdot \overline{\overline{K}}^* \cdot \overline{E}_0^*\right] d\tau} ; \qquad (15)$$

(b) Cutoff frequency of a waveguide

$$\omega - \omega_0 = \frac{-\omega\epsilon_0 \int_A \left[\overline{E}_0^* \cdot \overline{\overline{K}}_{op} \cdot \overline{E} - \overline{E} \cdot \overline{\overline{K}}^* \cdot \overline{E}_0^*\right] da}{\int_A \left[\mu_0 \overline{H}_0^* \cdot \overline{H} + \epsilon_0 \overline{E} \cdot \overline{K}^* \overline{E}_0^*\right] da} ; \qquad (16)$$

(c) Propagation constant of a waveguide ($e^{-\gamma z}$)

$$\gamma + \gamma_0^* = \frac{j\omega\epsilon_0 \int_A \left[\overline{E}_0^* \cdot \overline{\overline{K}}_{op} \cdot \overline{E} - \overline{E} \cdot \overline{\overline{K}}^* \cdot \overline{E}_0^*\right] da}{\int_A \overline{i}_z \cdot \left[\overline{E}_0^* \times \overline{H} + \overline{E} \times \overline{H}_0^*\right] da} \qquad (17)$$

Relations 15-17 are exact relations. Since the fields E and H are unknown, some approximations can be made if the perturbation is small. It is usually assumed (unless some more accurate approximation is available) that $\overline{E} \approx \overline{E}_0$, $\overline{H} \approx \overline{H}_0$. With this approximation, the denominator of (17) becomes $4P_z^{em}$, that is, four times the time-averaged power carried in the z-direction by the unperturbed wave.

b. Variational Principles

Since the dielectric-tensor operator $\overline{\overline{K}}_{op}$ is Hermitian when used with the boundary conditions (9)-(12), it is possible to derive variational principles for the characteristic quantities of the system.

(a) Resonant frequency of a cavity

$$\omega^2 = c^2 \; \frac{\int_\tau \left| \nabla \times \overline{E} \right| \, d\tau}{\int_\tau \overline{E}^* \cdot \overline{\overline{K}}_{op} \cdot \overline{E} d\tau} \; ; \tag{18}$$

(b) Cutoff frequency of a waveguide

$$\omega^2 = c^2 \; \frac{\int_A \left| \nabla_T \times \overline{E} \right| \, da}{\int_A \overline{E}^* \cdot \overline{\overline{K}}_{op} \cdot \overline{E} da} \; ; \tag{19}$$

(c) Propagation constant of a waveguide

$$\gamma = j\beta = \frac{\int_A \left[\overline{H}^* \cdot \nabla_T \times \overline{E} - \overline{E}^* \cdot \nabla_T \times \overline{H} + j\omega\mu_0 \left| \overline{H} \right|^2 + j\omega\epsilon_0 \overline{E}^* \cdot \overline{\overline{K}}_{op} \cdot \overline{E} \right] da}{2 \, \text{Re} \int_A \left[\overline{E} \times \overline{H}^* \right] \cdot \overline{i}_z da} \; . \tag{20}$$

This variational principle in Eq. 20 is valid only when $\gamma = j\beta$, and hence only in particular frequency regions.[5] The trial fields for the three variational principles must be continuous and differentiable, and must satisfy the boundary conditions.

Further details and applications may be found in Gardiol's thesis.[6]

F. Gardiol, A. Bers

References

1. P. E. Serafim and A. Bers, Interaction between an electron beam and plasmas, Quarterly Progress Report No. 70, Research Laboratory of Electronics, M. I. T., July 15, 1963, pp. 121-128.

2. W. P. Allis, S. J. Buchsbaum, and A. Bers, Waves in Anisotropic Plasmas (The M. I. T. Press, Cambridge, Mass., 1963).

3. D. L. Bobroff and H. A. Hauss, Uniqueness and Orthogonality of Small Signal Solutions in Electron Beams, Technical Report No. 31, Research Division, Raytheon Company, Waltham, Mass., 1958.

4. B. Lax and K. Button, Microwave Ferrites and Ferri-magnetics (McGraw-Hill Publishing Company, New York, 1962).

5. W. P. Allis, S. J. Buchsbaum, and A. Bers, op. cit., p. 176.

6. F. Gardiol, Approximate Techniques for Hot Plasma Waveguides and Resonators, S. M. Thesis, Department of Electrical Engineering, M. I. T., January 1965.

F. EMISSION PROCESSES BY AN OSCILLATOR MOVING AT SUPER-WAVE VELOCITY

When an oscillator is moving relative to a medium through which a certain kind of wave is propagating, it emits or absorbs a quantum, which is associated with the wave

field. The energy of an emitted quantum depends upon the relative velocity, as well as the characteristic frequency of an oscillator according to the Doppler effect. But if the relative velocity exceeds the phase velocity of the wave, qualitatively new phenomena appear. One of them has already been pointed out and discussed by Ginzburg, Fidman, and Zheleznyakov[1,2] as the "anomalous Doppler effect." The anomalous Doppler effect can be explained either classically or quantum mechanically. In this report other new phenomena will be discussed which have no classical analogues in this velocity region.

In the interaction between a single oscillator and a wave field if the diagonal elements of the interaction Hamiltonian between an oscillator and a wave field are zero with respect to the unperturbed eigenstates, the total momentum and energy are conserved in the emission or absorption processes.

$$\sum_\lambda n_\lambda \hbar k_\lambda + \overline{P} = const \tag{1}$$

$$\sum_\lambda n_\lambda \hbar \omega_\lambda + \frac{\overline{P}^2}{2M} + W = const, \tag{2}$$

where n_λ, $\overline{k}_\lambda$, ω_λ are the quantum number, wave vector, and frequency, respectively, of the wave mode specified by λ; $\overline{P}$, M, W are the momentum, the mass, and the internal energy, respectively, of the oscillator. Assuming that only one mode of the wave field is involved in the present process, and the wave vector is parallel to P, we have

$$W = const - \frac{1}{2M}(P-MU)^2, \tag{3}$$

where $U = \omega_\lambda/k_\lambda$ is the phase velocity. This is shown in Fig. XI-16.

This is a locus on which the momentum and the internal energy of the oscillator are confined during the interaction.

Let us consider only the emission processes, since the absorption occurs in the direction opposite to the emission on this locus. The internal energy of the oscillator is supposed to be changed by ΔW.

1. When the oscillator state is at 1 initially, it goes down the slope to 2 after emission of a quantum with a momentum $\hbar k_\lambda$, as required by Eq. 1. In this process W always decreases. This sort of process is called the "normal Doppler effect."

2. When the oscillator is moving slight faster than the phase velocity and located at 3, it can make a transition to 4 without losing its internal energy. This is a Cherenkov type of emission process.
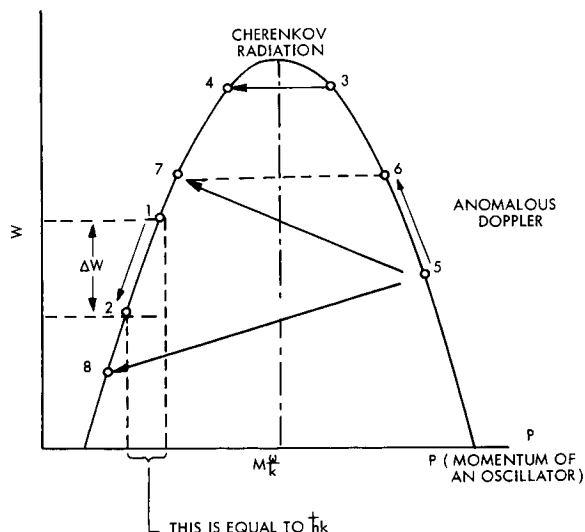
Fig. XI-16.   Diagram of transitions.

3.   When the oscillator velocity largely exceeds the phase velocity of the wave which is located at 5 it can make three different transitions.   A transition 5 → 6 is called the anomalous Doppler effect, in which the final velocity of the oscillator is always larger than U and the internal energy of the oscillator is always increased.   On the other hand, in a process 5 → 7, the oscillator increases the internal energy and at the same time is decelerated to a velocity smaller than U.   The other possibility is a process 5 → 8, in which the oscillator loses its internal energy as well as its kinetic energy.   In the last two processes the internal state of the oscillator goes through an energy tunnel to a final state, and we have no classical analogues for them.   A possible example will be found in the interaction of phonons with gyrating conduction electrons in a DC magnetic field. These processes occur spontaneously or through stimulation by the wave field.

T. Musha

References

1.   V. L. Ginzburg and V. Ya. Eidman, Soviet Phys. — USP 2, 874 (1960).

2.   V. L. Ginzburg, V. V. Zhalenznyakov, and V. Ya. Eidman, Phil. Mag. 7, 451 (1962).

G. ELECTRON CYCLOTRON RESONANCE DISCHARGE[*]

1. NEW MICROWAVE TECHNIQUE FOR MEASURING ELECTRON DENSITY

Since the last report[1] a new perturbation method of measuring the electron density of a plasma in a resonant cavity has been developed. The method is a variation of the method developed by Rose and Brown,[2] wherein the plasma shifts the resonant frequency of a cavity mode by changing the complex dielectric constant of the medium filling the cavity. The new method uses perturbations of many of the higher order cavity modes to measure the density of the plasma. The theory of the measurement is particularly simple.

Consider a resonant cavity partially filled with plasma which is excited at the resonant frequency of the nth higher order mode. Assume that there is no steady magnetic field, or, if one is present, that the ratio of the electron cyclotron frequency to the applied frequency is negligible compared with one. Then, according to Rose and Brown, the determining ratio $\Delta f/f$ of the nth mode as a result of the presence of the plasma is given by

$$\frac{\Delta f}{f} = \frac{\int_{V_p} \omega_p^2(\bar{r}) |\hat{e}_n(\bar{r})|^2 \, d^3\bar{r}}{2\omega^2 V_c}.$$

where $\omega_p(\bar{r})$ is the electron plasma frequency as a function of position, ω is the unperturbed resonant frequency of the nth mode, V_c and V_p are the cavity and plasma volumes, and $\hat{e}_n(\bar{r})$ is the eigen electric-field vector of the nth mode. Here the normalization

$$\int_{V_c} |\hat{e}_n(r)|^2 \, d^3\bar{r} = V_c$$

is assumed.

Now if the volume of the plasma V_p is large compared with a cubic half-wavelength at the resonant frequency of the nth mode, the integral can be evaluated as

$$\int_{V_p} \omega_p^2(\bar{r}) |\hat{e}_n(r)|^2 \, d^3\bar{r} = \omega_p^2 V_p,$$

where ω_p^2 is the average density of the plasma over the volume V_p. If Δf is considerably larger than the frequency separation between modes near the frequency f, the

---

plasma will cause many modes to sweep past a given frequency f. Thus if δf is the average mode separation, the total frequency shift Δf of the nth mode is given by

$$\Delta f \cong m \delta f,$$

where m is the number of modes that sweep past a given frequency f_o. The relation will be most accurate if m is much greater than one. Thus we are led to the relation

$$(\omega_p/\omega)^2 = -\frac{2m\delta f}{f}\frac{V_c}{V_p}.$$

Experiment

Two X-band waveguides were attached to the resonant box of the ECRD and microwave energy at approximately 9.5 Gc was coupled through the box and observed with a
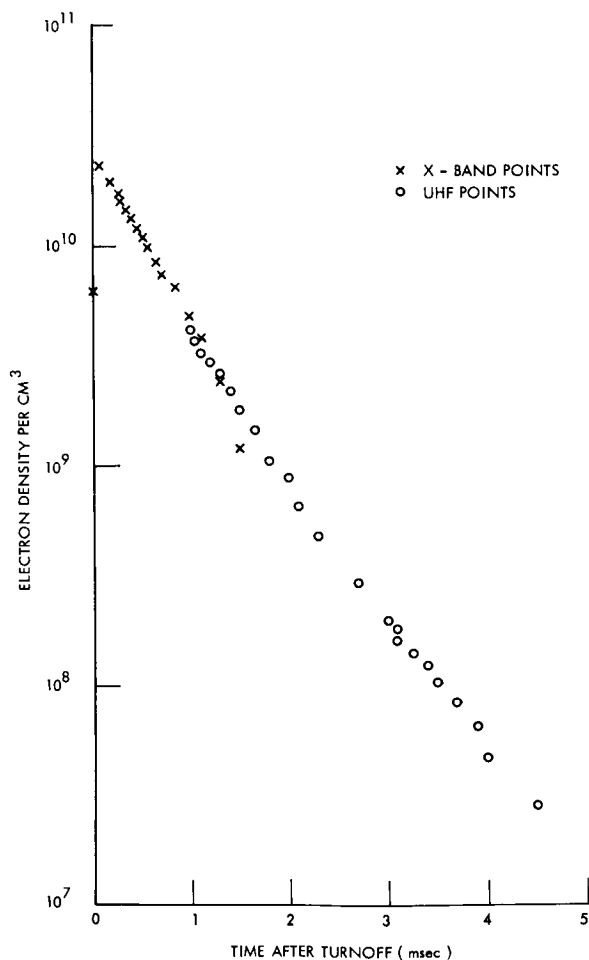

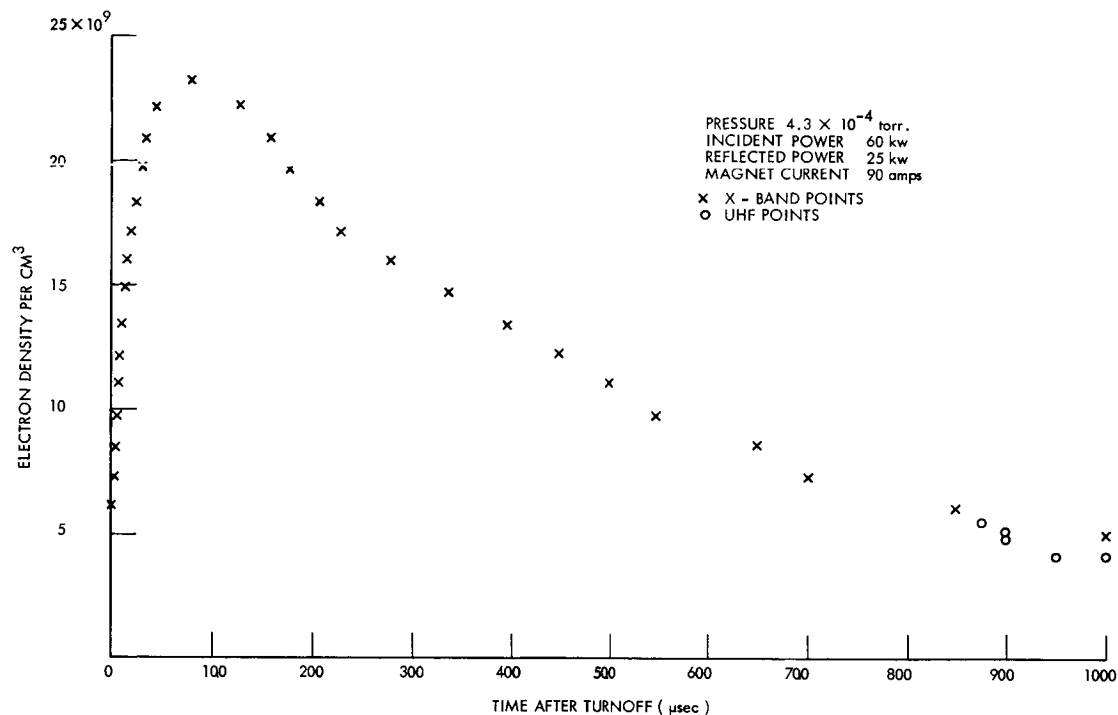Fig. XI-17. Plasma electron density between microwave pulses.

Fig. XI-18. Plasma electron density after turnoff measured with the X-band and UHF probing mode techniques. The plasma is assumed to lie in a cylinder 10 inches in diameter on the cavity axis.

crystal detector and an oscilloscope. As the frequency of the X-band energy was varied slightly, peaks in energy transmission were detected when different modes of the cavity were exactly at resonance. The mode number of these modes is approximately 10,000, and their separation was measured to be 1.69 mc/mode averaged over 100 modes, with a 10-mode maximum average of 1.98 mc/mode and a minimum of 1.27 mc/mode.

The discharge was run in the burst mode[1] and the plasma tuning of the X-band modes was observed in the afterglow after the last S-band power pulse. The parameter Δf was measured as a function of time by counting the number of modes after a given time which could be detected sweeping by the frequency 9.5 Gc. By using Eq. 1, the plasma density was calculated; some typical results are presented in Figs. XI-17 and XI-18. A comparison of results of this technique and of the UHF probing mode technique[1] is made in Fig. XI-18. In obtaining these points, all parameters of the system were carefully adjusted until they were reproducible on a pulse-to-pulse basis.

T. J. Fessenden

References

1. T. J. Fessenden, Quarterly Progress Report No. 76, Research Laboratory of Electronics, M. I. T., January 15, 1965, pp. 125-128.

2. D. J. Rose and S. C. Brown, J. Appl. Phys. 23, 1028-1032 (1952).

2. INTENSITY MEASUREMENTS OF THE ARGON 4880 Å LINE[*]

Intensity measurements of incoherent light emitted from an argon plasma have been made. The frequencies of particular interest were those used in argon lasers, specifically the 4880 Å and 5150 Å lines. The 5150 Å line, however, was never observed in the plasma studied here, and so the following results are for the 4880 Å line of ionized argon.

The argon plasma was created by an electron cyclotron discharge with approximately 60 watts of incident RF power at 2.8 kmc. The plasma was confined by a rectangular cavity and a symmetric magnetic mirror (see Fig. XI-19). The field strength ratio of
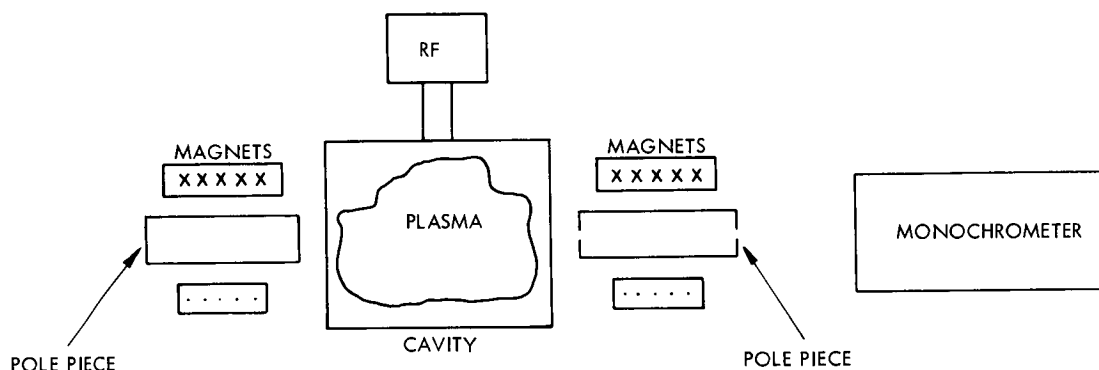

Fig. XI-19. Experimental arrangement.

the mirror was approximately 5:3. With this fixed ratio the field strength at the center of the cavity was varied from 500 gauss to 1000 gauss. The base pressure was approximately 5×10^{-6} mm Hg, and the discharge was studied at working pressures of 5×10^{-4}-5×10^{-5} mm Hg. A 3/4 inch hole was placed in one of the pole pieces and the light intensity emitted along the axial directions of the magnets was observed by a monochrometer set at 4880 Å. The output of the monochrometer was recorded for various values of the magnetic field and plasma density.

The monochrometer was then calibrated by a standard General Electric tungsten lamp. Using this calibration, we converted each plasma measurement to a value of n, which is the number of radiators per unit time per unit volume giving off a quantum of energy at 4880 Å. These results are shown in Figs. XI-20 through XI-22.
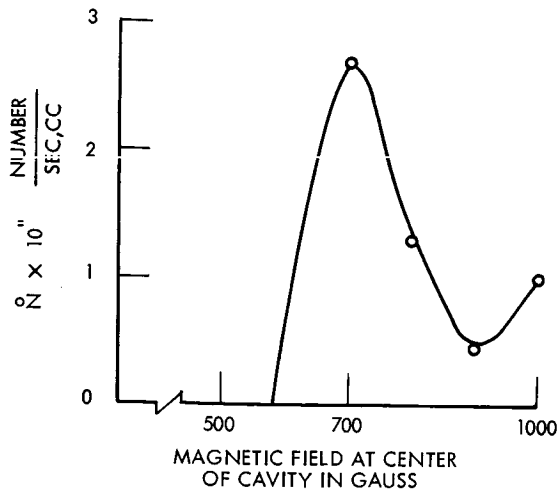

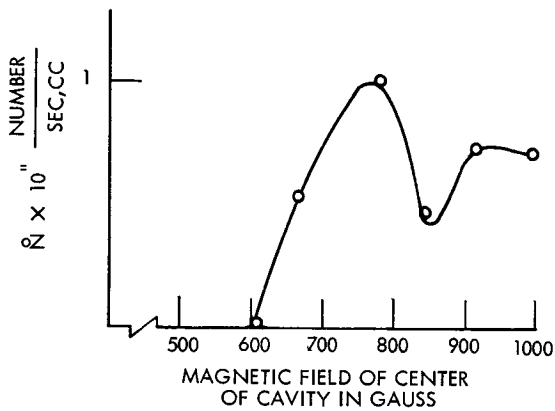Fig. XI-20. 4880 Å line argon intensity (5×10^{-4} mm Hg).

Fig. XI-21. 4880 Å line argon intensity (1×10^{-4} mm Hg).
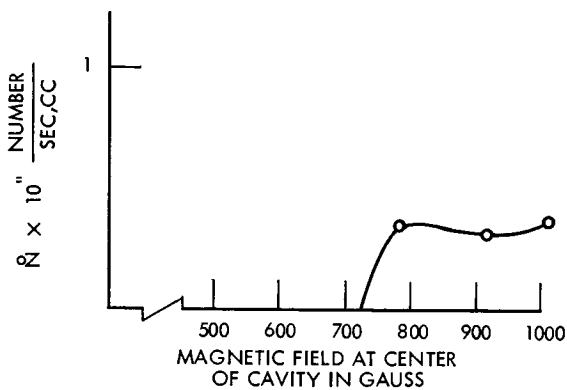

Fig. XI-22. 4880 Å line argon intensity (5×10^{-5} mm Hg).

We plan to continue this work and investigate the feasibility of pumping an argon laser with a cyclotron or beam-plasma discharge.

B. R. Kusse

H. FUSION BLANKET RESEARCH

With the completion of the doctoral thesis of L. M. Petrie, the initial phase of this work has been completed. The substantive content of five theses in this area of research, which have been submitted to the Department of Nuclear Engineering, M. I. T., in partial fulfillment of the requirements for the degrees indicated below, will be published in the Research Laboratory of Electronics technical report series. A list of authors and titles follows.

A. J. Impink, Neutron Economy in Fusion Reactor Blanket Assemblies

Ph. D. Thesis, January 1963, to appear as Technical Report 434.

W. G. Homeyer, Thermal and Chemical Aspects of the Thermonuclear
        Blanket Problem

Sc. D. Thesis, December 1962, to appear as Technical Report 435.

L. M. Lontai, A Study of a Thermonuclear Reactor Blanket with Fissile Nuclides

S. M. Thesis, May 1963, to appear as Technical Report 436.

P. S. Spangler, Fusion Reactor Blanket Experiment

Sc. D. Thesis, February 1965, to appear as Technical Report 437.

L. M. Petrie, Gamma Ray Spectra in Fusion Blanket Mock-ups

Sc. D. Thesis, March 1965, to appear as Technical Report 438.

D. J. Rose

I. NONADIABATIC DIFFUSION IN TOROIDAL GEOMETRY

An apparatus is being built to produce a circulating electron current in toroidal geometry by nonadiabatic injection of a cw electron beam. Although partly motivated by the problem of cw injection into closed geometry, our major objective is a study of the effect of small nonadiabatic perturbations on particle diffusion in space and velocity. A circulating beam is an ideal medium for these studies because of the simple relations between diffusion coefficients and beam lifetime.

1. Nonadiabatic Scattering

The motion of charged particles is said to be adiabatic when the magnetic fields change slowly enough along the particle trajectory. In this case, both μ, the magnetic moment, and J, the longitudinal invariant, may be treated as constants of the motion. The particle trajectory is then well described by the guiding-center approximation which predicts, except for electrical and gravitational drifts, particle motion along surfaces of fixed μ and J. If the adiabatic conditions are violated, the adiabatic invariants suffer a random fluctuation with a resultant diffusion. The nonadiabatic motion depends, as the adiabatic motion does not, on the detailed relation between the particle's phase in its motion around the guiding center and the shape and extent of the field perturbations.

Several theoretical analyses of this process have appeared[1,2] but no definitive experimental work has been done.

2. Experiment

The device will be a race track with 6 m circumferential length, and 10-cm minor diameter (see Fig. XI-23). The electron injection energy will be 4 kev to yield a 2-cm
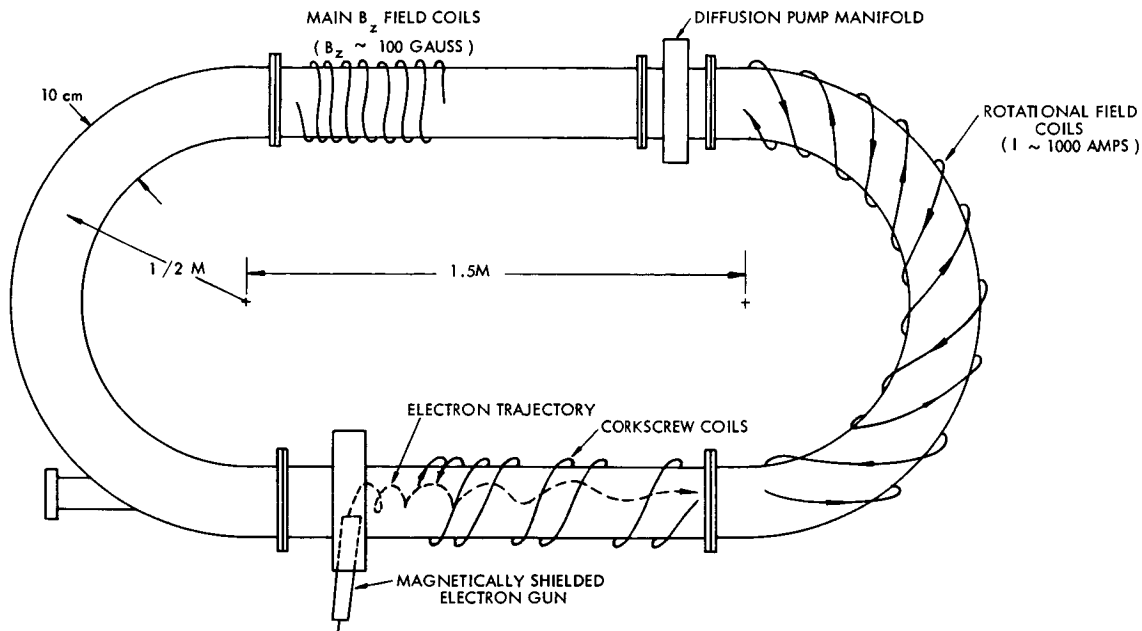
Fig. XI-23. Toroidal magnetic trap.

Larmor radius with the planned main field of 100 gauss. The electrons will be injected essentially perpendicularly to the field with only enough $v_{\parallel}$ to miss the gun snout after one orbit. They will then enter an accelerating corkscrew[3,4] which, by cumulative non-adiabatic perturbations, will increase $v_{\parallel}$ at the expense of $v_{\perp}$, so that in one pass the motion will be almost entirely along B.

The injected electrons would undergo severe curvature drifts in the U bends of the torus if no corrective action were taken. The drift distance for a particle of Larmor radius r_b is πr_b for complete traversal of the bend. In our system, because of the nature of the corkscrew injection scheme, the electron Larmor orbit is comparable to the minor diameter of the torus, and, therefore, this amount of uncorrected drift would be intolerable. Hence, we shall make use of helical windings to impart a rotational transform to the magnetic field lines. These differ from the windings found on Stellarator-type devices in that we cannot rely on drift cancellations over many transits

of the entire system but must cancel drifts within a single U bend. To study these complicated particle trajectories, we have developed a computer program to integrate the equations of motion and to determine the magnetic field surfaces. Figure XI-24 shows the motion of a particle in a U bend with a 2π rotational transform. The important
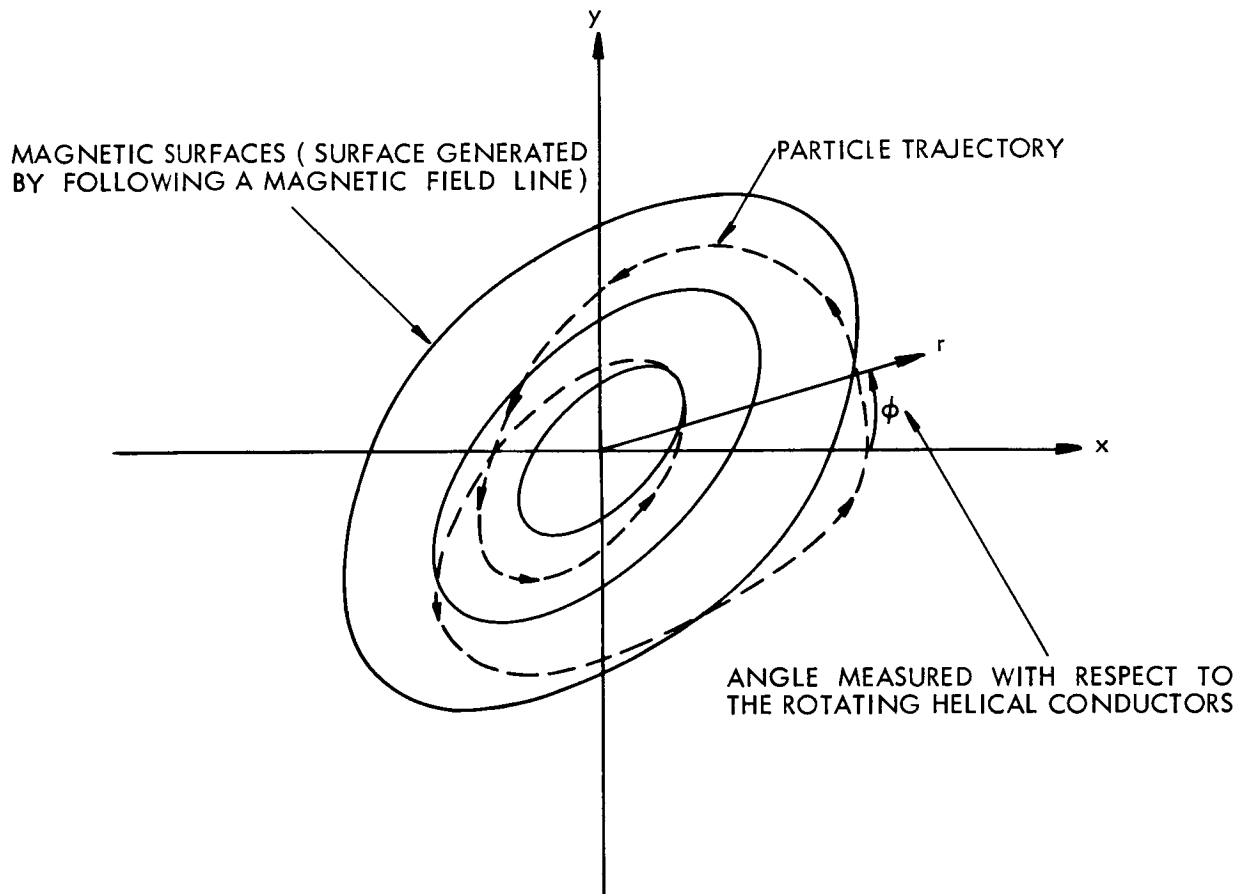
Fig. XI-24. Two dimensional projection of a trajectory and magnetic surfaces with a rotational transform of 2π.

quantity is the radial component of the vector connecting the particle position at the beginning and end of its transit through the U bend. Figure XI-25 shows the effect of many transits through a U bend, illustrating the order of magnitude of the remaining spatial diffusion effects for one particular choice of helical transform strength. Our aim is to make particle losses caused by imperfections in the drift cancelling scheme small compared with those resulting from nonadiabatic losses from the corkscrew or other intentionally introduced perturbations that are to be studied.

$B_r = Ar \sin(\ell\theta - h_z)$

$B_\theta = Ar \cos(\ell\theta - h_z)$

$B_z = 1.0$

$A - 0.15$

$h = 0.314$

$\ell = 2$

$\tau = 0$

$\dot{x} = \dot{y} = 0$

$\dot{z} = 2.0$

LENGTH OF U BAND = 120 cm
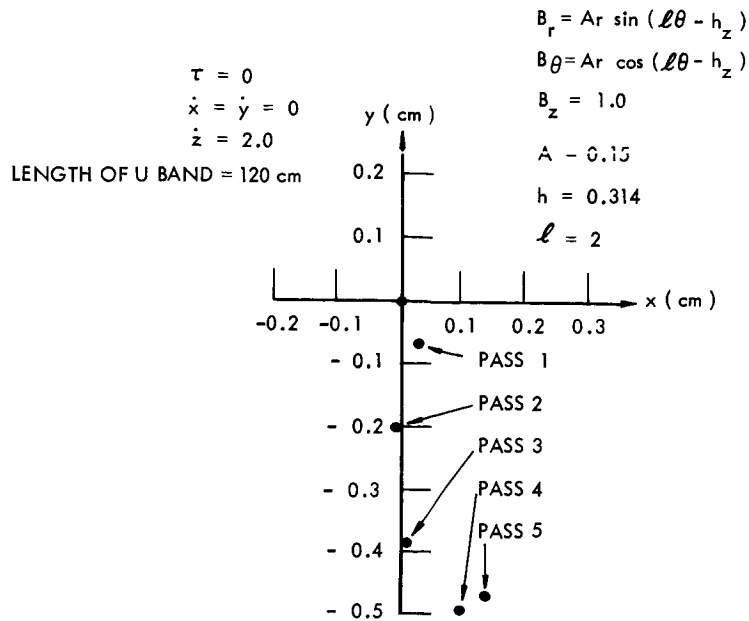
Fig. XI-25. Cumulative plot of residual drifts after multiple passes through a U bend with rotational transform of 2π at r = .5 cm.

Principal diagnostic methods will involve Rogowsky coils for detection of the modulated circulating beam and flourescent screens for visual location of beam position.

R. W. Moir, L. M. Lidsky

References

1. R. C. Wingerson, T. H. Dupree, and D. J. Rose, Phys. Fluids 7, 1475 (1964).

2. A. F. Laing and B. M. Robson, J. Nucl. Energy, Part C, 3, 146(1961).

3. L. M. Lidsky, Phys. Fluids 7, 1484 (1964).

4. C. Kapetanakos, S. M. Thesis, Department of Nuclear Engineering, M. I. T., August 1964.

J. PLASMA TURBULENCE STUDIES

The construction of the Hollow-Cathode Discharge III device described in Quarterly Progress Report No. 76 (pages 130-133) has been completed. The magnet coils have been tested successfully and maximum field in excess of 4 Kgauss can be obtained over the region of interest. The profile of the axial field on the centerline of the system and at the wall of the vacuum chamber is shown in Fig. XI-26.
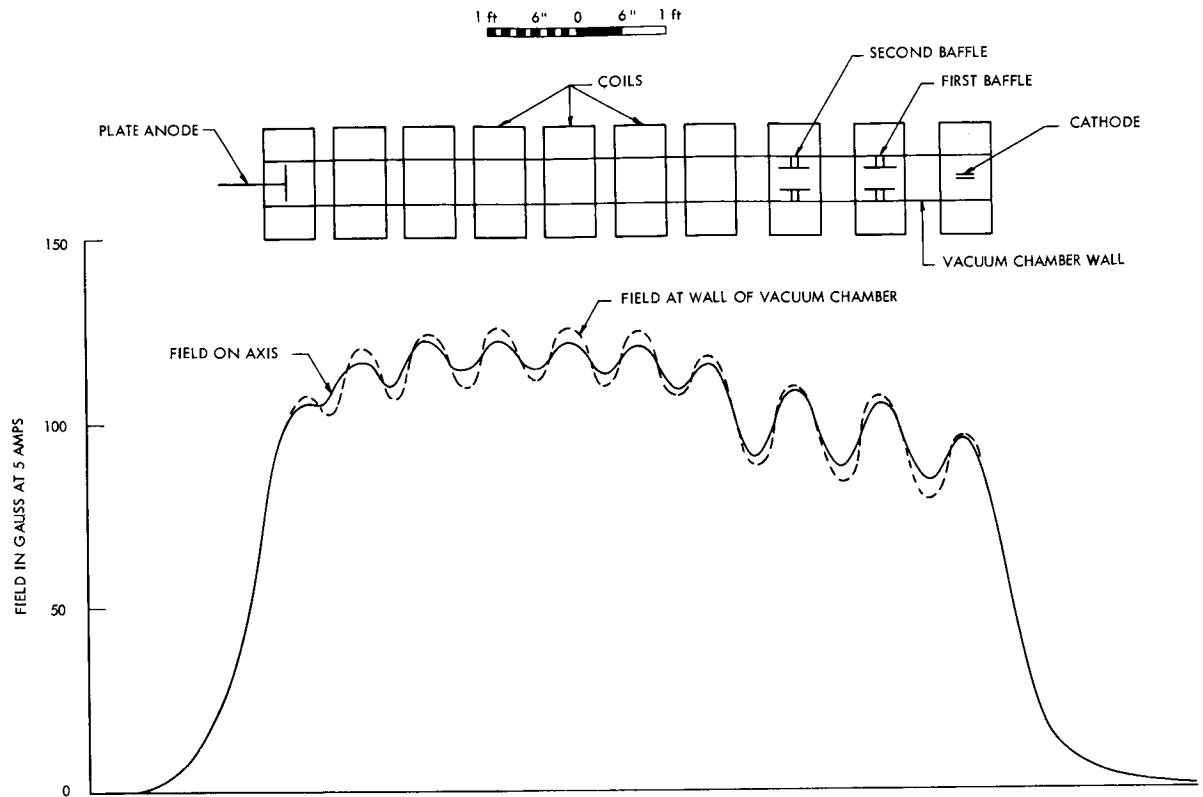
Fig. XI-26.   Profile of axial field of HCD III system.

Shims are now being installed to minimize the variations in the axial field, after which the device will be operational.   The first phase of experiment will be the determination of operating parameters of the plasma;   the next phase will be the turbulence studies.

J. C. Woo

K.   STUFFED-CUSP PLASMA FACILITY

Construction has begun on a "stuffed-cusp" magnetic trap.   This is a "minimum B" system consisting of a simple spindle cusp with an axial conductor.   The axial current creates a system with a nonzero minimum of induction which should enhance the single-particle (adiabatic) containment properties of the cusp.

We propose to generate a plasma inside the device by injecting an electron beam through the line cusp and making use of the beam-plasma interaction.   In this manner, we hope to generate a hot electron plasma with densities of the order of $10^{12}/cm^3$.   If this scheme fails to work, microwave equipment is available to create a plasma by electron-cyclotron resonance heating.   The resonant frequency then can be adjusted so
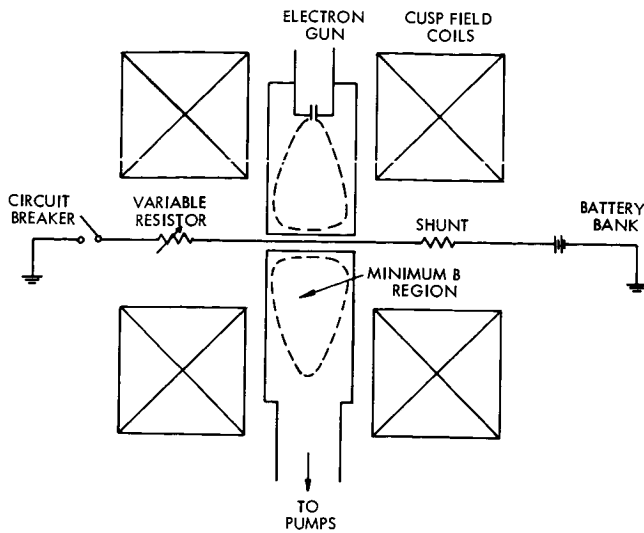
Fig. XI-27. Schematic diagram of the stuffed cusp.

that the heating region coincides with the center of the containment volume.

A schematic diagram of the facility is shown in Fig. XI-27. The cusp field will be generated by two existing tape-wound coils driven by a 60-kw generator. Axial current will be furnished by a bank of 12-v storage batteries which will give 14. 7 kA through the water-cooled copper central conductor. A circuit breaker will be constructed employing a dielectric slab driven between movable electrodes. Interchangeable resistor links consisting of stainless-steel and copper tubes will be used to make fine adjustments in current.

The resulting magnetic field configuration, computed by using the MAFCO code,[1] is plotted in Fig. XI-28. The largest closed magnetic isobar is 1.16 kgauss, while the minimum induction is 0.62 kgauss; this gives a mirror ratio of 1.9.

A cylindrical stainless-steel vacuum can (24-inch diameter X 7 inches) has been constructed with 7 radial ports. A removable vacuum wall has been built around the central conductor so that the conductor may be withdrawn and the system run as a simple cusp. A 4-inch oil-diffusion pump and baffle that give a pumping speed of 300 liters/sec have been installed.
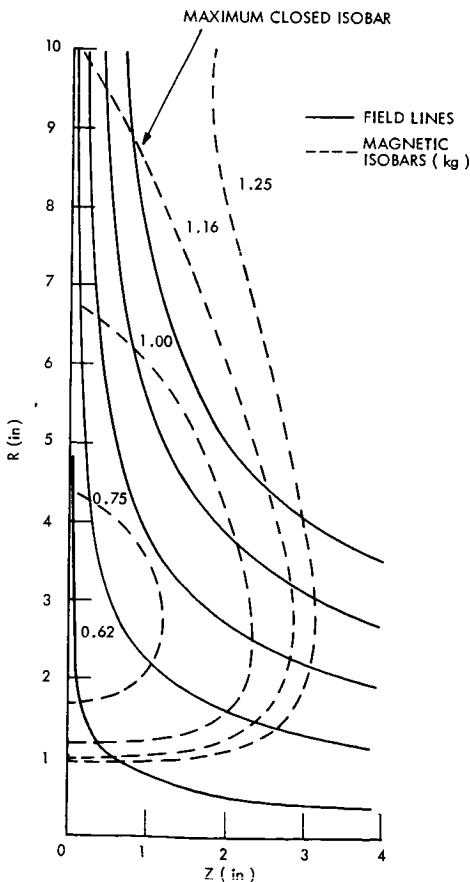
Fig. XI-28. Magnetic field plot.

Proposed diagnostic techniques for the facility are microwave interferometry; x-ray, light, and RF spectral measurements; and single-particle detection with scintillation or solid-state counters.

We propose to study the effect of a nonzero minimum of the induction on the stability and containment properties of a cusp. Also, studies of the beam-plasma interaction or of electron-cyclotron heating in this geometry may be performed.

C. E. Wagner, L. M. Lidsky

References

1. The MAFCO code was developed at Lawrence Radiation Laboratory, Livermore, California.

L. MOTION IN NONADIABATIC FIELDS[*]

Let us consider a charged particle moving in a magnetic field which has a characteristic length of the order of the particle's Larmor radius. The particle's magnetic moment will no longer be conserved; as a result, an external beam can be trapped between magnetic mirrors. The phenomenon is widely recognized and has formed the basis of many publications.[1-4] In this report we attempt to obtain further insight into the interaction by studying the equation of motion,

$$\frac{d\vec{v}}{dt} = \vec{v} \times \vec{\omega}; \qquad \vec{\omega} = \frac{e\vec{B}}{m} \qquad (1)$$

in curvilinear coordinates running along the field lines. The coordinate system is obtained by defining the unit vectors $\hat{t}$, $\hat{n}$, and $\hat{b}$, where $\hat{t}$ is along the field direction, n is in the direction of the principal normal, that is, in the direction of the instantaneous radius of curvature, and $\hat{b}$ is in the direction of the bi-normal, that is, the direction in which the field lines twist. These vectors are related by the Frenet-Seret formulas

$$\frac{d\hat{t}}{ds} = \kappa \hat{n}$$

$$\frac{d\hat{n}}{ds} = -(\kappa\hat{t}+\tau\hat{b}) \qquad (2)$$

$$\frac{d\hat{b}}{ds} = \tau \hat{n}$$

where κ is the inverse radius of curvature, τ is a measure of the torsion, and s is a measure of arc length along the field lines. As a result of employing these relations and the definition

$$ds = v_t dt \qquad (3)$$

the equation of motion becomes

$$\frac{dv_t}{ds} = \kappa v_n \qquad (4a)$$

$$\frac{dv_n}{ds} = -\kappa v_t + \left(\frac{\omega}{v_t} - \tau\right) v_b \qquad (4b)$$

$$\frac{dv_b}{ds} = \qquad -\left(\frac{\omega}{v_t} - \tau\right) v_n. \qquad (4c)$$

These equations can be simplified by defining $v_t = v_n + jv_b$ and combining (4b) and (4c) to obtain

$$\frac{dv_+}{ds} = -\kappa v_t - j(\beta - \tau) \, v_+ \tag{5a}$$

$$\frac{dv_t}{ds} = \kappa v_n, \tag{5b}$$

where $\beta = \omega / v_t$. For a weak nonadiabatic perturbation, the change in the particle's magnetic moment per transit will be small unless some resonance occurs. The obvious approximation is to assume v_t to be constant. Then (5a) and (5b) uncouple, and we can treat κ, τ, and β as known functions that specify the magnetic field. The physical meaning of (5a) is clear if we convert from a space to a time description,

$$\frac{dv_+}{dt} + j(\omega - v_t \tau) \, v_+ = -\frac{v_t^2}{R}. \tag{6}$$

Since $\kappa = 1/R$, we see that the force that changes the magnitude of the perpendicular velocity is due to the centripetal acceleration experienced when the particle attempts to follow the perturbed field lines. The phase of the particle's rotation is specified by the imaginary term, and this is affected by both the magnitude of the field and its torsion.

A direct integration gives

$$\Delta v_+ = -\int_0^L \kappa v_t \exp -j \left\{ \int_s^L (\beta - \tau) \, ds' \right\} ds, \tag{7}$$

which is the desired result, the change in the perpendicular velocity resulting from a transit through the perturbation.

The physical content of Eq. 7 is most easily visualized by comparison with the formal solution of the equation of radiative transfer:

$$\Delta I = \int_0^L \eta \exp - \left\{ \int_s^L a \, ds' \right\} ds, \tag{8}$$

where I is radiative flux, η is the emissivity, and a is the absorption per unit length. The total change of flux in a beam traversing an active medium is given by summing, at the end point, the contribution of the emission at each point on the path diminished by the absorption of that quantum over the remaining path length. Referring to Eq. 7, we see that the effect of the perturbation can be interpreted in an analogous manner. The change in v_+ produced by the curvature at each point on the particle's trajectory is

added at the end of the perturbation with an appropriate phase determined by the factor $(\beta-\tau)$ which is the net rotation per unit path length.

Let us now consider the particular example of the stellarator type of field which, near the axis, can be represented by

$$\omega_r = \omega_0 h(kr)^{\ell-1} \sin (\ell\theta-kz)$$

$$\omega_\theta = \omega_0 h(kr)^{\ell-1} \cos (\ell\theta-kz) \tag{9}$$

$$\omega_z = \omega_0 (1+h(kr)^{\ell} \cos (\ell\theta-kz)).$$

Here θ is the particle's azimuthal position, and h is a smallness parameter. We assume that (kr) is also small and approximately constant over the particle orbit. Neglecting terms in h^2 compared with h, we find

$$\kappa = h(kr)^{\ell-1} \left\{ k - (\ell-1) \frac{\partial\theta}{\partial z} \right\} \tag{10a}$$

$$\tau = \left\{ k - (\ell-1) \frac{\partial\theta}{\partial z} \right\}. \tag{10b}$$

If we consider a particle moving along a field line with only a small perpendicular velocity, we can assume that θ is a constant. Then (7) gives

$$\Delta v_+ = -h(kr)^{\ell-1} kv_t \int_0^L \exp -j \left\{ (\beta-k)\, z + \left(\frac{\beta}{k}\right) h(kr)^{\ell} \sin (\ell\theta-kz) \right\} \, dz. \tag{11}$$

The resonant nature of the interaction is evident from the $(\beta-k)$ term in the exponent. This measures the phase of the particle relative to the field. Because the field curvature is constant, the particle sees a force of constant magnitude and if it rotates with the field $(\beta=k)$, its phase relation to this force is preserved. Therefore a large transfer can occur between its parallel and perpendicular kinetic energy. The second term in the exponent is of higher order and represents the change in phase caused by a variation in the magnitude of the particle's cyclotron period as the main field fluctuates.

We consider now a bumpy mirror type of field given by

$$\omega_z = \omega_0 (1+h \sin kz)$$
$$\omega_r = -\frac{(kr)}{2} \omega_0 h \cos kz. \tag{12}$$

Employing the same assumptions as before, we find

$$\kappa = kh\left(\frac{kr}{2}\right) \sin kz \tag{13}$$

$$\tau = 0$$

and the interaction is described by

$$\Delta v_+ = -kh\left(\frac{kr}{2}\right) v_t \int_0^L \sin kz \, \exp{-j}\left\{\beta z - \left(\frac{\beta}{k}\right) h \cos kz\right\} dz. \tag{14}$$

We see that now the basic resonance is caused by a variation of the field curvature along the particle's trajectory. Thus we can see a basic difference in these two interactions. In the case of stellarator fields the particle experiences a constant centripetal force, and it is the twist of the field lines which determines the resonance by adding the velocity changes in phase. In the bumpy-mirror case the phase propagation is independent of the perturbation, but the curvature force fluctuates in space so that the velocity changes from different parts of the orbit will add in phase at the end point.

This manner of thinking suggests a method of attack for minimizing the velocity change for trapped particles. Although κ and τ are not independent, a consideration of their different natures suggests that their effects may be made to cancel for specific particle velocities. If this velocity is chosen near the loss cone, the possibility exists of reducing the scattering of particles past that velocity by the perturbation. The ultimate particle lifetime would then be determined by other mechanisms.

In addition to its function as a guide to an intuitive understanding of nonadiabatic interactions, Eq. 7 can also be used to explain the complicated subresonances observed in some recent trapping experiments.[5] These subresonances appear when it becomes necessary to discard the assumption that θ is a constant. This occurs when the particle acquires enough perpendicular velocity so that its Larmor orbit becomes comparable with the characteristic lengths of the field. In the experiment performed by Demirkhanov and his co-workers,[5] a constant pitch $\ell = 3$ stellarator field was used. Rewriting (11) with the full form of (10b), we obtain

$$\Delta v_+ = \int_0^L \kappa v_t \, \exp{-j}\left\{(\beta_0 - k)z + (\ell - 1)\theta + \left(\frac{\beta}{k}\right) h(kr)^\ell \sin(\ell\theta_0 - kz)\right\} dz, \tag{15}$$

where the z dependence of θ in the first-order term has been ignored for convenience. For off-axis particles we can write

$$\theta(z) = \Delta\theta \sin\beta_0 z, \tag{16}$$

where $\Delta\theta \leq \pi$. Then if we employ the identity

$$e^{j\rho \sin \phi} = \sum_{n=-\infty}^{\infty} J_n(\rho) \, \ell^{jn\phi}, \tag{17}$$

Eq. 15 becomes

$$\Delta v_+ = \sum_{a,b} \int_0^L \kappa v_t J_a\left(\frac{\beta}{k} h(kr)^\ell\right) J_b((\ell-1)\Delta\theta) \, \exp -j\left\{(\beta_o(b+1)-k)z+a(\ell\theta_o-kz)\right\} \, dz. \tag{18}$$

To lowest order, $J_a = \delta_{ao}$ and resonances appear at

$$k = (b+1)\beta_o \tag{19}$$

with their relative amplitudes given by the ratio of the Bessel functions $J_b((\ell-1)\Delta\theta)$. As h increases, subresonances should appear at

$$(a+1)k = (b+1)\beta_o. \tag{20}$$

Two well-defined subresonances have been observed for particles injected along the axis, where h is quite small, and their spacing is given by (19) with $b = 0, 1$. Furthermore, if $\Delta\theta$ is taken equal to π for these particles (because their orbits intersect the axis), the relative amplitudes of the observed resonant peaks are given quite closely by the ratio

$$\frac{\Delta I_o}{\Delta I_1} = \left(\frac{\Delta v_{\perp o}}{\Delta v_{\perp 1}}\right)^2 \approx \left(\frac{J_o(2\pi)}{J_1(2\pi)}\right)^2. \tag{21}$$

When the beam was injected off axis, which is equivalent to an increase in h, a complicated fine structure appeared, superimposed on the main peaks. This can be qualitatively identified as the effect of the resonance condition given by Eq. 20. Thus it is apparent that a great deal of information can be extracted from the equations of motion when they are written in the form of Eq. 7. Although the agreement with experiment indicated in this report is mainly qualitative, more exact solutions can be obtained by using iterative techniques.

J. F. Clarke

References

1. K. D. Snelnikov et al, Soviet Phys. — Tech. Phys. 5, 236 (1960A).

2. R. C. Wingerson, Phys. Rev. Letters 6, 446 (1961).

3. E. W. Laing and A. E. Robson, J. Nucl. Energy (Part C. Plasma Physics) 3, 146 (1961).

4. R. C. Wingerson, T. H. Dupree, and D. J. Rose, Phys. Fluids 7, 1475 (1964).

5. R. A. Demirkhanov, Yu. V. Kursanov, D. C. Baratov, and C. V. Klarin, Soviet Phys. — Tech. Phys. 9, 45 (1964).

M. DETERMINING THE ELECTRON DISTRIBUTION FUNCTION FROM SCATTERED LIGHT II[*]

The electron velocity distribution function in a plasma can be uniquely related to the spectral intensity of photons scattered out of an incident, collimated, monochromatic beam. Thus, in principle, the velocity distribution function can be completely determined. A previously reported initial investigation[1] into the relation between the spectrum and the electron distribution function (including its sensitivity) has been extended and concluded. Details of the work have been presented in an S. M. thesis[2]; only the principal results are given in this report.

The electron velocity distribution function, $f(\underline{\beta})$, where $\underline{\beta}(=\underline{v}/c)$ is the velocity of the electrons, is mapped into the spectral intensity $s(\underline{n}\omega)$ by the operator L_0. $s(\underline{n}\omega)$, the power per unit frequency interval at ω scattered into a unit solid angle in the direction of the unit vector $\underline{n}$, is given by

$$s(\underline{n}\omega) = L_0 f(\underline{\beta}).$$

(1)

Similarly, the inverse relation is

$$f(\underline{\beta}) = L_0^{-1} s(\underline{n}, \omega).$$

(2)

The operators L_0 and L_0^{-1} have been derived[2] for both relativistic and nonrelativistic cases. Since the relativistic expressions are too complex to be very useful, only the nonrelativistic expressions will be given here:

$$L_0 = \frac{3}{8\pi} \sigma_T n_e I_0 \frac{\omega^2}{\omega_0^3} \left(1-(\underline{n}\cdot\underline{n}_0)^2\right) \int d\underline{\beta}\ \delta(g(\underline{n}, \omega, \beta))$$

(3)

and

$$L_0^{-1} = -\frac{12\omega_0^3}{\sigma_T I_0 n_e} \int \frac{d\Omega_{\underline{n}}}{4\pi} \frac{(1-\underline{n}\cdot\underline{n}_0)}{(1-(\underline{n}\cdot\underline{E})^2)(1-\underline{n}\cdot\underline{\beta})^3} \frac{\partial^2}{\partial\omega^2}\Bigg|_{\frac{\omega}{\omega_0} = \frac{1-\underline{n}_0\cdot\underline{\beta}}{1-\underline{n}\cdot\underline{\beta}}}$$

(4)

The notation is the same as that used in a previous report.[1] In L_0^{-1} the integration is over all directions of $\underline{n}$, and the derivative must be evaluated at the given point. Equations 3 and 4 show that the relationship between the distribution function and the spectral intensity of scattered light is unique.

Table XI-1. Magnitude, shape functions, and half-widths of various distributions.

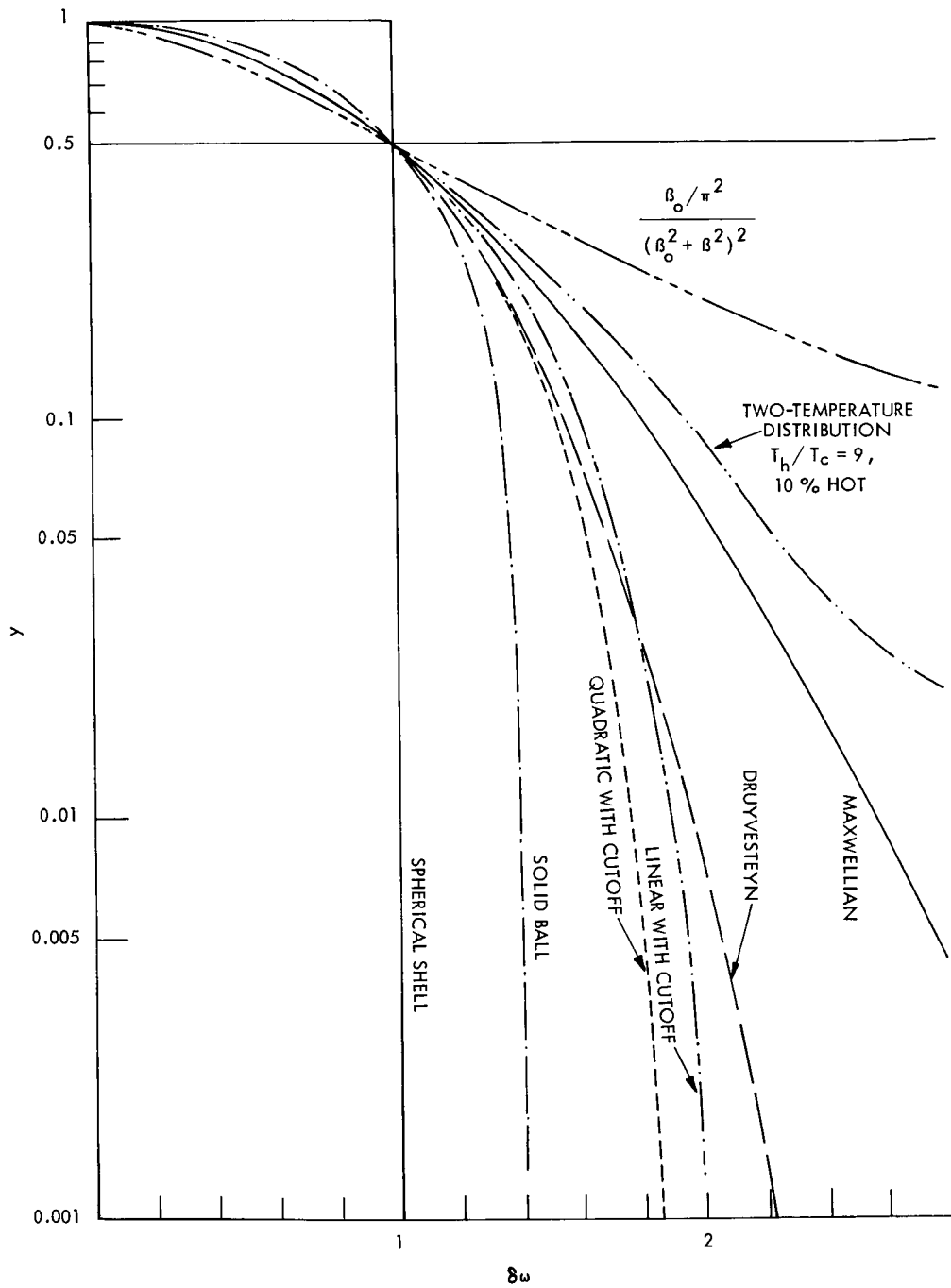| f	$\dfrac{S_m}{n_e r_e^2 I_0 \dfrac{\omega^2}{\omega_0^2}\dfrac{\sin^2\theta}{\sqrt{2}\,\omega_0\beta_0\sqrt{1-\cos\phi}}}$	y	$\dfrac{\delta\omega}{\sqrt{2}\,\omega_0\beta_0\sqrt{1-\cos\phi}}$	$\dfrac{\beta_{avg}}{\beta_0}$	$\dfrac{(\beta^2)_{avg}}{\beta_0^2}$
$\dfrac{1}{\pi^{3/2}\beta_0^3}\exp\left[-(\beta/\beta_0)^2\right]$	$\dfrac{1}{\sqrt{\pi}}$	$\exp\left[-\left(\dfrac{\omega-\omega_0}{\delta\omega}\right)^2\right]\ln 2$	$\sqrt{\ln 2}=.832$	$\dfrac{2}{\sqrt{\pi}}$	$\dfrac{3}{2}$
$\dfrac{3\exp\left[-(\beta/\beta_0)^4\right]}{4\pi\Gamma\left(\frac{7}{4}\right)\beta_0^3}$	$\dfrac{3\sqrt{\pi}}{8\Gamma(7/4)}$	$1-erf\left[\left(\dfrac{\omega-\omega_0}{\delta\omega}\right)^2.474\right]$	.688	$\dfrac{3}{4\Gamma\left(\frac{7}{4}\right)}=.816$	$\dfrac{3\Gamma\left(\frac{5}{4}\right)}{4\Gamma\left(\frac{7}{4}\right)}=.74$
$\dfrac{\beta_0/\pi^2}{\left(\beta_0^2+\beta^2\right)^2}$	$\dfrac{1}{\pi}$	$\dfrac{1}{1+\left(\dfrac{\omega-\omega_0}{\delta\omega}\right)^2}$	1	—	—
$\dfrac{3}{\pi\beta_0^3}\left(1-\dfrac{\beta}{\beta_0}\right)$	1	$1-\dfrac{3}{4}\left(\dfrac{\omega-\omega_0}{\delta\omega}\right)^2+\dfrac{1}{4}\left(\dfrac{\omega-\omega_0}{\delta\omega}\right)^3$	$1/2$	.6	.4
$\dfrac{1-u(\beta/\beta_0)}{\frac{4}{3}\pi\beta_0^3}$	$\dfrac{3}{4}$	$1-\dfrac{1}{2}\left(\dfrac{\omega-\omega_0}{\delta\omega}\right)^2$	$\dfrac{1}{\sqrt{2}}$	$\dfrac{3}{4}$	$\dfrac{3}{5}$
$\dfrac{15\left(1-(\beta/\beta_0)^2\right)}{8\pi\beta_0^3}$	$\dfrac{15}{16}$	$1-(2-\sqrt{2})\left(\dfrac{\omega-\omega_0}{\delta\omega}\right)^2+\left(\dfrac{2-\sqrt{2}}{2}\right)^2\left(\dfrac{\omega-\omega_0}{\delta\omega}\right)^4$	.541	$\dfrac{5}{8}$	$\dfrac{3}{7}$
$\dfrac{\delta(\beta-\beta_0)}{4\pi\beta^2}$	$\dfrac{1}{2}$	$1-u(\omega\pm\delta\omega)$	1	1	1
$\dfrac{S}{\pi^{3/2}\beta_0^3}\exp\left[-(\beta/\beta_0)^2\right]$ $+\dfrac{F}{\pi^{3/2}\beta_f^3}\exp\left[-(\beta/\beta_f)^2\right]$	$\dfrac{S\beta_f+F\beta_0}{\sqrt{\pi}\,\beta_f}$	$\dfrac{\beta_f\beta_0}{S\beta_f+F\beta_0}\left\{S\exp\left[-\left(\dfrac{\omega-\omega_0}{\Delta\omega_s}\right)^2\right]\Delta\omega_s+F\exp\left[-\left(\dfrac{\omega-\omega_0}{\Delta\omega_f}\right)^2\right]\Delta\omega_f\right\}$	—	$\dfrac{2}{\sqrt{\pi}}\left(S+\dfrac{\beta_f}{\beta_0}F\right)$	$\dfrac{3}{2}\left(S+\left(\dfrac{\beta_f}{\beta_0}\right)^2F\right)$

Fig. XI-29.   Plot of the shape functions.

From (4) it can be seen that to determine an arbitrary velocity distribution function at a point in velocity space, a great deal of information about the spectral intensity of the scattered light is required. In particular, a specific average, over all directions of scattering, of the second derivative of the spectral intensity (evaluated at a specific frequency which depends on the direction of scattering) is required. In the present simpler experimental arrangements, light is scattered by plasma electrons out of a laser beam and detected in only one or two directions. What is needed in addition to Eq. 2 is a more pragmatic approach, one which would be useful to an experimentalist. Such an approach can be developed by examining Eq. 1.

From (1) it has been shown[1] that the spectrum observed in any particular direction is equivalent to an analysis of the distribution function in one dimension along a specific direction. Thus, one can expect to carry out experimentally only a finite number of one-dimensional density analyses in velocity space of the distribution function. These one-dimensional analyses are useful insofar as some kind of symmetry can be assumed. Even if some kind of symmetry is assumed, the second derivative of the spectrum (a very difficult quantity to determine experimentally) must be measured in order to ascertain the causative distribution function. A method of avoiding such a difficult measurement is to compare the experimental spectrum with spectra calculated by using known distribution functions, that is, to construct a library. A unique method of comparing the experimental and calculated spectra is given below. A rudimentary library calculated from some spherically symmetric distribution functions is illustrated in Fig. XI-29 and the corresponding parameters are listed in Table XI-1.

In order to compare the calculated and experimental spectra, we must note that the shape of the spectrum is the same as the one-dimensional shape of the distribution function and hence the magnitude of the spectrum (corresponding to the electron density) and the width of the spectrum (corresponding to the 1-D velocity spread of the electrons) can be removed as separate factors, leaving only the shape of the spectrum.

This remaining spectral shape is uniquely related only to the one-dimensional distribution function shape. Variations in the two remaining dimensions of the distribution function are indeterminate. The magnitude and width of the spectrum are removed by scaling the plot of the spectrum in such a manner that the ordinate of the maximum is unity, and the abscissa of the half-maximum is unity. The result is a shape function, y. Comparison of the experimental shape functions with the library shape functions (Fig. XI-29) then gives the type of distribution function. The magnitude of the half-width at half-height, $\delta\omega$, gives the average speed and energy (Table XI-1), and the absolute magnitude of the spectral intensity at its maximum, s_m, gives the electron density n_e (Table XI-1).

The accuracy of this analysis depends, of course, on how precisely the experimental spectrum is known. Although the relationship is unique in the one-dimensional sense,

it is well to note that for most distribution functions the relative difference in the resulting spectrum is greatest in the "tail" where the magnitude of the experimental signal is minute, most difficult to measure, and easily obscured by noise. For example, there is little difference between the shape functions of the linearly decreasing and Druyvesteyn distribution functions until the Doppler frequency shift is greater than two half-widths (that is, $2\delta\omega$). Deviations from the commonly assumed Maxwellian distribution can be measured, however, and the average speed and energy can easily be obtained.

T. S. Brown

References

1. T. S. Brown, Determining the electron distribution function from scattered light, Quarterly Progress Report No. 74, Research Laboratory of Electronics, M. I. T., July 15, 1964, pp. 146-149.

2. T. S. Brown, S. M. Thesis, Department of Nuclear Engineering, M. I. T., September 10, 1964.

N. THOMSON SCATTERING DIAGNOSTICS OF A HOLLOW-CATHODE ARC PLASMA[*]

Preliminary results of an experiment designed to study Thomson scattering of laser radiation as a plasma diagnostic method were reported in Quarterly Progress Report No. 75 (pages 81-84). This work continues and data yielding a much higher signal-to-noise ratio have been obtained. The higher signal to noise was obtained by using an input system of short focal length for the laser radiation which focused a large fraction of the laser output energy into the interaction volume. The use of this system eliminated the possibility of doing the small-angle scattering experiment mentioned in the previous report. It appeared, however, that this would be a very difficult experiment with the present ruby laser for several other reasons. We decided, therefore, to concentrate the effort on obtaining good large-angle scattering data.

The plasma electron temperatures and densities obtained from the present scattering data generally agree well with those obtained from pulsed Langmuir probe measurements performed by M. Lubin. Lubin's measurements are described in Section XI-O. In general, the Thomson scattering data points lie on a Gaussian distribution with departures at certain shifted wavelengths. One of these departures occurs at wavelengths very near the incident laser wavelength. This anomaly was apparently the source of the "low temperature" distribution reported in Quarterly Progress Report No. 75. The reasons for these deviations from a Gaussian distribution are not understood at present.

1. Hollow-Cathode Arc Plasma Source

The hollow-cathode arc facility used for the scattering experiments was not described in any detail in Quarterly Progress Report No. 75. We shall discuss its properties now. The apparatus as viewed looking toward the laser is shown in Fig. XI-30. The cathode is a 3-inch length of 1/8 inch diameter, 0.010-inch wall, tantalum tubing. The anode is a 1-inch I.D., water-cooled hollow structure through which gas may be fed. A low-conductance pumping baffle was installed between the cathode chamber and the scattering chamber with each chamber being separately pumped. This allowed a relatively low background pressure of the order of 10^{-4} torr in the scattering chamber, even with relatively high gas flow rates through the cathode. The anode was grounded with the cathode running at a negative potential. The addition of anode flow seemed to stabilize the arc in this configuration and provide increased electron density. Without the anode gas flow the arc seemed to have difficulty negotiating the pumping baffle. When the baffle was allowed to float electrically it attained a potential near that of the cathode. When it was
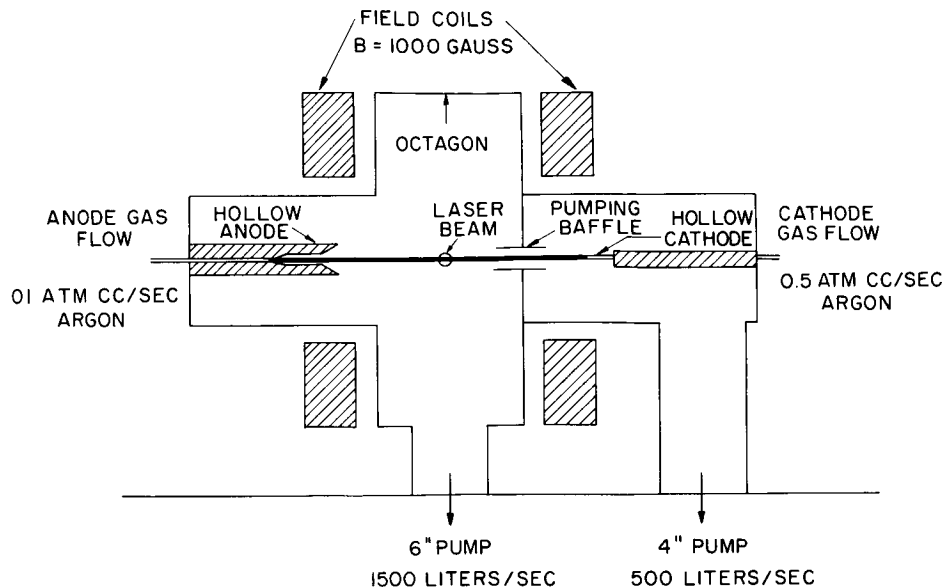
Fig. XI-30. Hollow-cathode arc plasma facility.

grounded approximately 70 per cent of the current went to the baffle, and only 30 per cent to the anode. The arc column in the scattering chamber was very diffuse.

By adding a very small gas flow through the anode (small compared with that through the cathode), the baffle could be made to float at zero potential and thus draw no current when grounded. In this way, the baffle could be eliminated, as far as its effect on the electrical characteristics of the arc are concerned, and thus serve only as a pumping baffle. The arc column in the scattering chamber under these conditions was a well-defined bright column, approximately 1 cm in diameter, with an electron density of approximately $5 \times 10^{13}/cm^3$ at an arc current of 40 amps. The total arc length was 17 inches. The field coils provided a magnetic field of approximately 1000 gauss at the cathode and anode, which were placed at the mirror points, and a field of approximately 500 gauss at the center in the scattering chamber.

Additional flexibility was attained with the introduction of an anode gas feed, in that mixed gas operation could be accomplished. With the cathode running in argon, hydrogen or helium could be introduced through the anode. The result was a relatively pure hydrogen or helium plasma in the scattering chamber, made possible by the strong differential pumping between the two chambers. Of course, some argon was present, but the measured intensities of major argon lines were more than a factor of 20 less than the intensities observed with the arc running in pure argon. The visual effect was quite striking. A bright blue argon arc would disappear into the pumping baffle from the cathode side and emerge as a bright red hydrogen arc in the scattering chamber. The arc would also be run in pure helium and by replacing the tantalum cathode with a

piece of Tungsten-Rhenium alloy tubing of the same dimensions, operation in pure hydrogen was possible.

2. Thomson Scattering Experiment

A schematic view of the scattering apparatus as seen from the cathode end of the arc is shown in Fig. XI-31. The output of the 100-joule ruby laser is focused into the scattering volume by the 50-cm focal length quartz input lens. The focal point for parallel light is at the scattering center, while the focal point for all light leaving the end of the ruby rod is just in front of the entrance to the offset conical-beam dump. This is the advantage of this short focal length system over the old 180-cm focal length system. All light from the laser is focused near the scattering center, while with the old system only the light emitted in the plane-parallel mode from the laser was focused at all, since the laser rod was inside the focal distance of the lens. The offset conical-beam dump was generally of the same type as that used by Thompson[1] in his experiments, except that the cone was offset to eliminate scattering from the necessarily finite-sized tip.
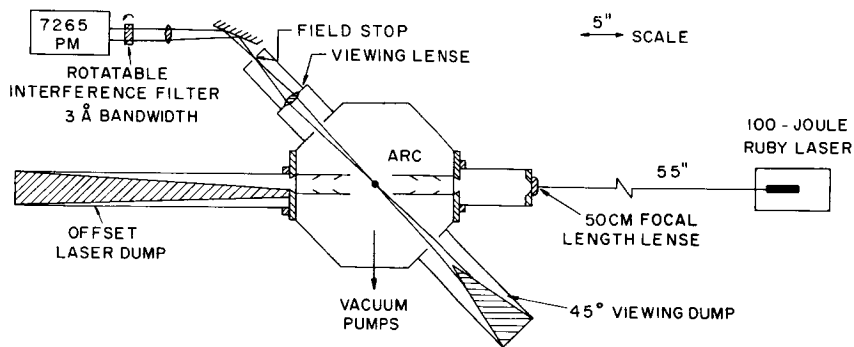

Fig. XI-31.   45° scattering arrangement with 50-cm focal-length lens viewed from cathode end.

Scattering was observed at 45°, the scattering volume being defined by the intersection of the laser beam with the image of the field stop formed by the viewing lens. For the plasmas under study in this device, 45° was a sufficiently large angle so that coherent effects could be neglected (see Quarterly Progress Report No. 75, pp. 81-84). The spectrum of scattered radiation in this case is just the Doppler-broadened spectrum mirroring the electron velocity distribution function in one dimension. After passing through the field stop, the scattered radiation was made parallel by a second lens before passing through the interference filter. The spectrum was scanned by rotating the 3 Å bandwidth interference filter. The radiation was detected by an RCA 7265 photomultiplier with an S-20 photocathode.

Laser light, scattered from the walls of the apparatus, was important only near the

laser line, and disappeared rapidly as the filter was tuned away from the line. The light baffling and collimation was sufficiently good so that this noise was only one-third of the total signal at the laser wavelength. The largest source of noise on the signals was the natural Bremsstrahlung from the plasma. Since the light from the plasma was continuous, this noise could be reduced by using a 150-μsec time constant integrator at the oscilloscope input; 150 μsec was short enough compared with the 1-msec laser pulse duration to preserve the laser pulse shape. The residual noise ultimately limited the magnitude of the smallest observable signal. All data points presented are the result of an overlay of three successive oscilloscope traces taken under the same conditions. The error bars indicate the vertical width of the overlaid traces. The scattered radiation was observed to be fully plane-polarized, as was the laser output. The scattered radiation was synchronous with the input laser radiation. No delayed effects such as those reported by Thompson[2] were observed. Such effects could be produced by raising the neutral background pressure in the scattering chamber by throttling the pump.
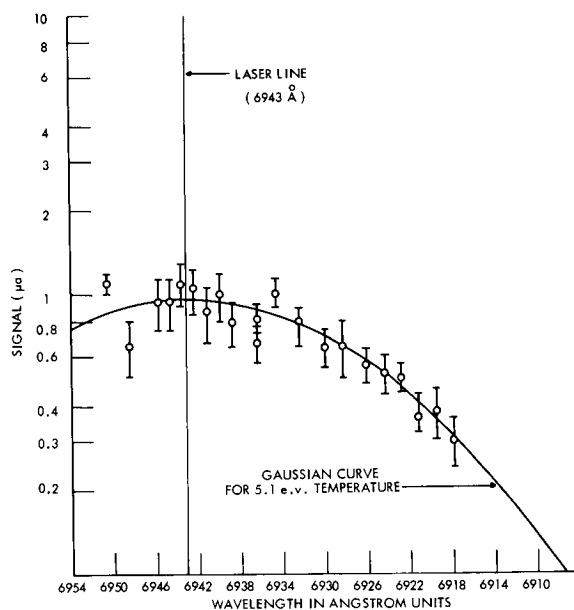
Fig. XI-32.   Thomson scattering signal vs wavelength. (Wavelength variation by rotating 6950.5 Å filter.)

Figure XI-32 exhibits some typical data plotted against wavelength, obtained with the arc running in argon. These data were obtained with a 6950.5 Å filter, so that some points are available on both sides of the laser line. A Gaussian curve corresponding to an electron temperature of 5.1 ev is drawn for comparison. Note the departure from the Gaussian, occurring 6-8 angstrom units either side of the laser line.

Figure XI-33 exhibits the same data as in Fig. XI-32 replotted against electron energy. The energy of the electron responsible for the scattering at a given shifted wavelength is proportional to the wavelength shift squared. A Gaussian electron velocity distribution is represented by a straight line on this plot.
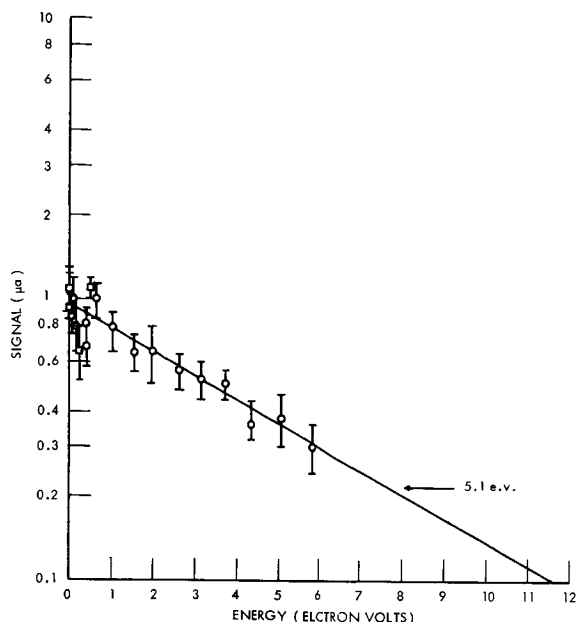

Fig. XI-33. Thomson scattering signal vs electron energy.
(Wavelength variation by rotating 6950.5 Å filter.)
Φ Points from long-wavelength side.
Φ Points from short-wavelength side.

Data obtained with a 6938 Å filter and with the arc running under the same conditions as for Figs. XI-32 and XI-33 are shown in Fig. XI-34. Note, again, the apparent departures from the Gaussian line drawn on the figure. The reasons for these departures from a Gaussian distribution are not at present understood. They are possibly due to actual deviations of the electron velocity distribution from Maxwellian. More likely, they are due to effects from electronically excited atoms or ions such as Raman scattering, or Rayleigh scattering from nearby emission lines.

Agreement with the probe data of M. Lubin, described in Section XI-O, is generally good. For an arc current of 40 amps in argon, he obtains electron temperatures in the range of 6.5-8.5 ev. We obtain temperatures in the range 4.5-5.5 ev, somewhat lower. This discrepancy is in the direction expected. Lubin finds electron densities in the range 4.5-7.5 × 10^{13}. Calibrating the system by using Rayleigh scattering from a known density of nitrogen, we obtain densities of 4 × 10^{13} which can be considered good
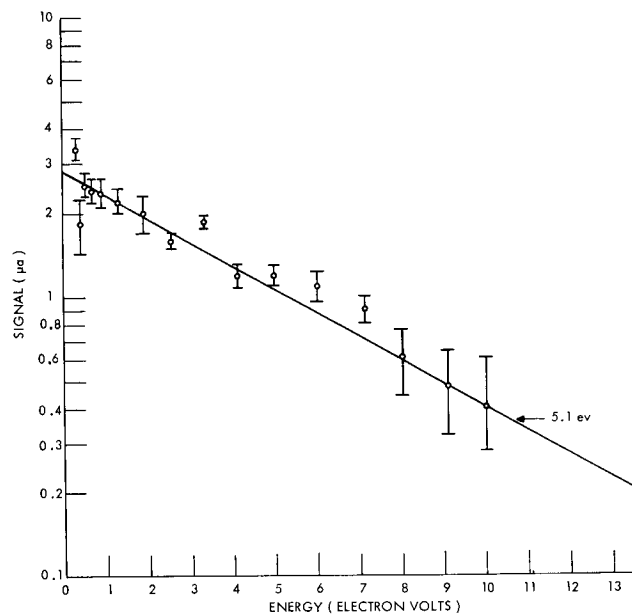
Fig. XI-34. Thomson scattering signal vs electron energy.
(Wavelength variation by rotating 6938.2 Å filter.)

agreement, since the uncertainties involved in obtaining electron densities from probe measurements are large.

The discrepancy between the electron temperature obtained from the probe data and the scattering data can probably be understood in the following way. The scattering system is sensitive to electron velocities only in the plane of scattering, which in this case is perpendicular to the magnetic field, so that with this system we can measure a perpendicular temperature $T_\perp$. The probe, however, collects electrons from an oval-shaped region, with the long dimension of the oval along the field. Thus the probe is primarily sensitive to parallel electron velocities, $T_\parallel$, with some contribution from T. Since the electrons gain their energy from the electric field of the arc, which is parallel to the magnetic field, perpendicular energy is obtained only through relaxation by collision. At the electron densities and temperatures under consideration here the mean-free path for an electron is only slightly smaller than the anode-cathode distance. Thus it might be expected that $T_\perp$ would be less than $T_\parallel$.

Further evidence for this reasoning can be obtained by examining the electron temperature as a function of magnetic field strength for both the probe and scattering measurements. The probe data shown in Fig. XI-41 in Section XI-O (Lubin's report) exhibit a tendency of constant or slightly decreasing measured electron temperature with increasing magnetic field. On the other hand, the Thomson scattering data, shown in Fig. XI-35 exhibit a tendency of increasing the electron temperature with increasing
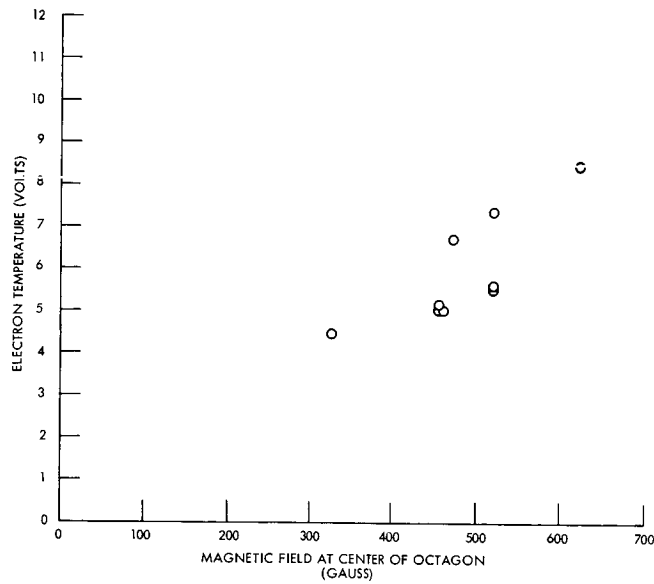
Fig. XI-35.   Thomson scattering data. (Electron temperature
vs magnetic field strength.)

magnetic field. This would be expected on the basis of the present discussion. The diameter of the arc column is observed to decrease as the magnetic field is increased, with the result that a higher ion density is exhibited in Fig. XI-40 in Section XI-O. A higher ion density reduces the time for relaxation of parallel energy to perpendicular energy, thereby yielding closer agreement between $T_\perp$ and $T_\parallel$.

Scattering data were also obtained with the arc running in hydrogen and helium but no unusual effects were observed.

E. T. Gerry

References

1. E. Thompson and G. Fiocco, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M. I. T., April 15, 1963, pp. 74-80.

2. E. Thompson, Quarterly Progress Report No. 75, Research Laboratory of Electronics, M. I. T., October 15, 1964, pp. 68-80.

O. PULSED LANGMUIR PROBE MEASUREMENTS ON A HOLLOW-
CATHODE DISCHARGE[*]

Langmuir probe measurements have been made on the hollow-cathode discharge
described in Section XI-N.

In order to avoid melting probes in the high-density plasma, characteristics were
obtained by a transient technique by using a simple passive circuit. The circuit is
diagrammed in Fig. XI-36. The probe, when idle, is at floating potential. When the
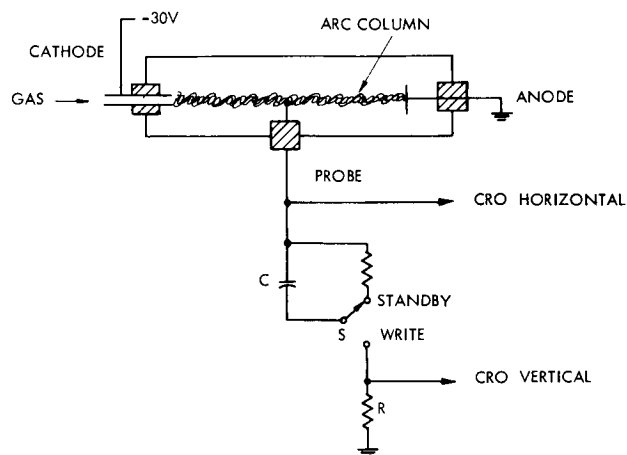
Fig. XI-36.   Circuit concept.

switch is thrown to the write position, the probe is momentarily placed at anode
potential, and the plasma electron current charges the capacitor C through R. The
probe voltage and the voltage across R are displayed as the horizontal and vertical
deflections, respectively, of an oscilloscope trace. When the capacitor is charged,
the probe is once again at floating potential and draws no current. Ion cur-
rent at potentials more negative than floating potential is measured by the usual
DC methods.

Typical records of probe characteristics and their logarithmic rectifications are
shown in Fig. XI-37. Electron temperature and ion density are calculated according
to the formulas of Bohm, Burhop, and Massey.[2]

$I - I_s$ (AMPS)

10
8
6
4

2

1
0.8
0.6
0.4

0.2

0.1

SLOPE $\}$ = 7 VOLTS
T_e

ION DENSITY ≈ 7 X 10^{13}/cc
.075" DIAMETER SPHERICAL PROBE
ON PLASMA AXIS
ION SATURATION CURRENT ≈ 300 ma

T_e = 3.7 VOLTS
ION DENSITY ≈ 2.8 × 10^{13}
1/8" DIAMETER DISC PROBE
APPROXIMATELY 1 4" FROM PLASMA
AXIS
I_s ≈ 60 ma

0   -2   -4   -6   -8   -10
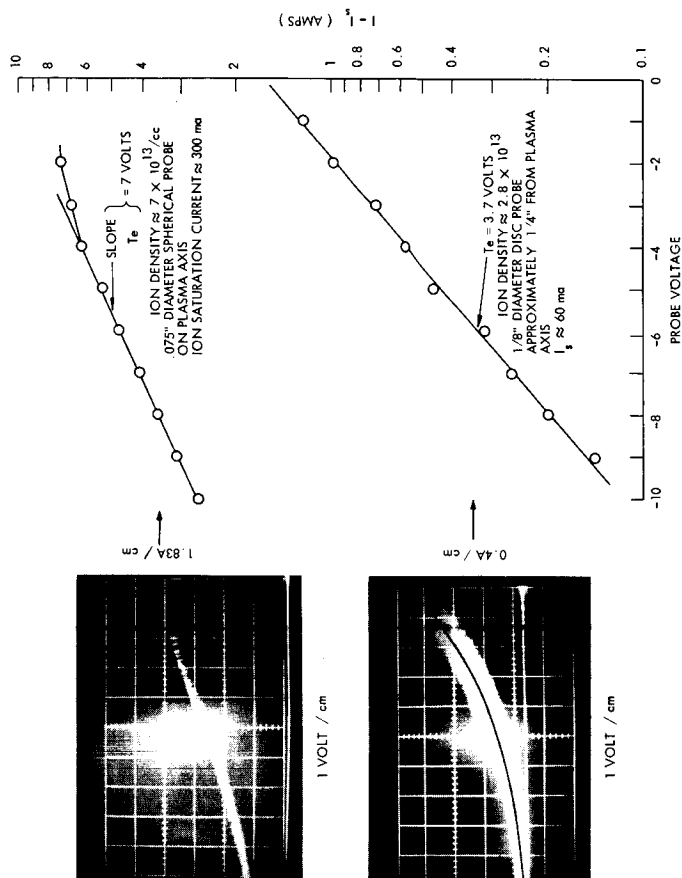
PROBE VOLTAGE

1.83A / cm

0.4A / cm

1 VOLT / cm

1 VOLT / cm

Fig. XI-37.   Records of probe characteristics and their logarithmic rectifications.

Plots of electron temperature and ion density against arc current and axial magnetic field measured at various times with various probes are shown in Figs. XI-38 through XI-41. Measurements were made on the axis of the plasma, in the position from which Thomson scattering was observed by E. T. Gerry. All probe measurements were made on argon plasmas.
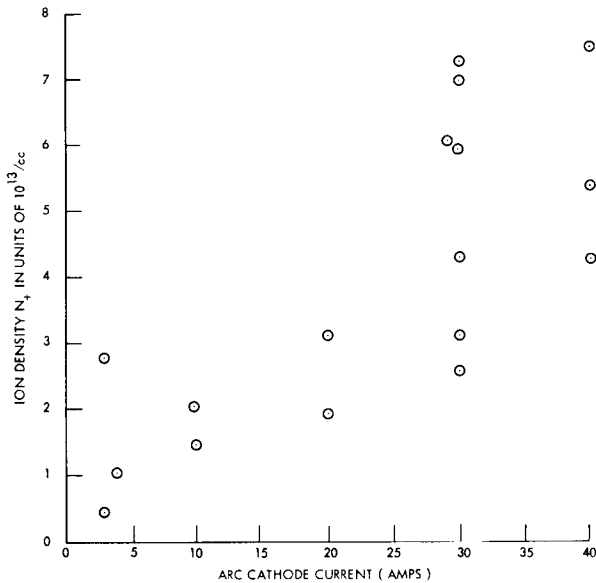

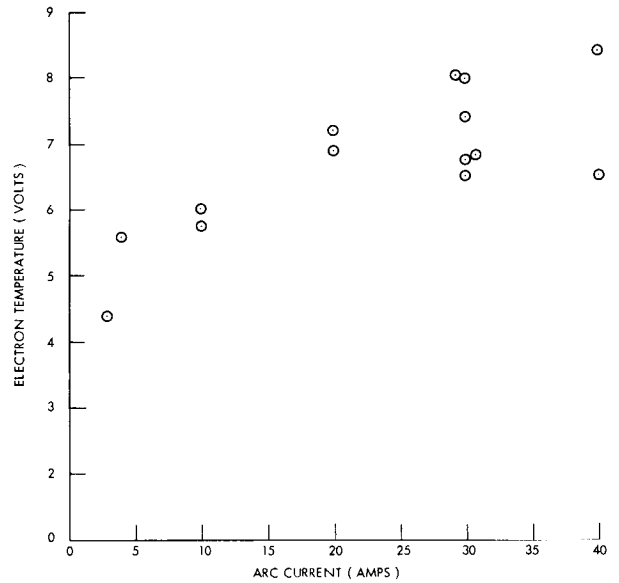Fig. XI-38. Ion density vs arc current. Magnetic field, 450 gauss; probe centered in the arc.

Fig. XI-39. Electron temperature vs arc current. Magnetic field, 450 gauss; probe centered in plasma.
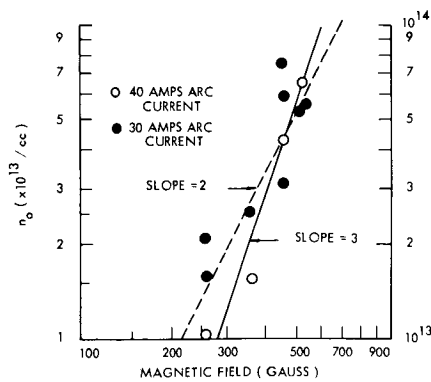
Fig. XI-40. Ion density vs magnetic field.
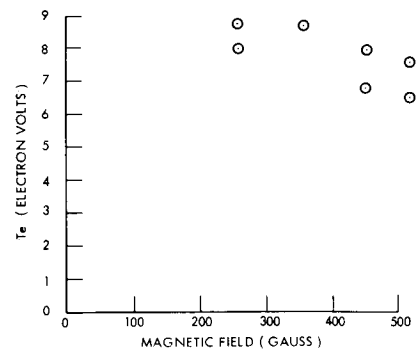
Fig. XI-41. Electron temperature vs magnetic field (30 amps arc current).

Table XI-2 gives a summary of results obtained by using probes of various shapes and sizes at one of Gerry's standard conditions. The comparison of these data with the Thomson scattering data has been reported in Section XI-N.

Table XI-2. Summary of results.

Gas Argon

Cathode flow   0.62 atm-cc/sec

Anode flow   0.175 atm-cc/sec

Magnetic Field   454 gauss

| | Arc Current (amps) | | | |
	10	20	30	40
Ion Density (units of 10^{13}/cc)	1.5-2	2-3.3	3-7	4.5-7.5
kTe (electron volts)	6	7	6.5-8	6.5-8.5

M. D. Lubin

References

1. E. T. Gerry, Research Laboratory of Electronics, M.I.T., Quarterly Progress Report No. 75, October 15, 1964, p. 81.

2. D. Bohm, E. H. S. Burhop, and H. S. W. Massey, The Use of Probes for Plasma Exploration in Strong Magnetic Fields. The Characteristics of Electrical Discharges in Magnetic Fields, edited by A. Guthrie and R. K. Wakerling (McGraw-Hill Book Company, New York, 1949), Chapter II.

P. ELECTRON TRANSPORT IN THREE-COMPONENT PLASMAS

1. Introduction

The purpose of this report is to present a technique for analysis of electron transport phenomena in three-component plasmas.

Transport phenomena in nonuniform gaseous mixtures have been analyzed by others.[1-4] In particular, Chapman and Cowling[4] have developed a formalism for the solution of a set of Boltzmann equations through a series of successive approximations. This formalism has been successfully used in the field of gas dynamics, under the assumption that, to a first approximation, all components of the mixture have the same temperature.

A modified Chapman-Cowling approach is used here to determine the electron distribution function and current and heat flux in a three-component plasma in which the electron temperature is different from the ion and neutral-particle temperatures. A similar method has recently been presented by Stachanov and Stepanov.[3] These authors, however, treated the charged-particle collisions by means of the small-angle Landau approximation and used a hard-sphere model for electron-neutral collisions. As well as removing these restrictions on the collision integrals, the present analysis yields the electron transport parameters in a form that is more amenable to physical interpretation.

This report has two main parts. First, a perturbation method similar to Chapman and Cowling's will be used to solve the Boltzmann equation for electrons. Then, the practicality of the perturbation method will be illustrated through its application to both reference Lorentz plasmas and a three-component plasma.

2. Perturbation Solution of the Boltzmann Equation

a. The Boltzmann Equations

The Boltzmann equations in a steady state, three-component plasma consisting of electrons (e), ions (i), and neutrals (n), may be written

$$\overline{v}_\alpha \cdot \nabla_r f_\alpha + \frac{q_\alpha \overline{E}}{m_\alpha} \cdot \nabla_v f_\alpha = \sum_\beta J_{\alpha\beta}(f_\alpha, f_\beta); \quad \alpha, \beta = e, i, n, \tag{1}$$

where $J_{\alpha\beta}$ is the collision integral for collisions between species α and β. This integral may be written in Boltzmann form as

$$J_{\alpha\beta}(f_\alpha, f_\beta) \equiv \iiint (f'_\alpha f'_\beta - f_\alpha f_\beta) \, g_{\alpha\beta} b \frac{db}{dx} \, dx d\epsilon d\underline{v}_\beta \tag{2}$$

where $g_{\alpha\beta} = |\bar{v}_\alpha - \bar{v}_\beta|$; b is the impact parameter; x and ϵ are the polar and azimuthal angles describing the rotation (in the center of mass coordinates) of the relative velocity vector during the collision; and $f'_\alpha = f_\alpha(\bar{r}, \bar{v}'_\alpha)$, with $\bar{v}'_\alpha$ the velocity before the collision. In writing Eq. 2 for like-particle collisions $(J_{\alpha\alpha})$, the second subscript is omitted. This eliminates confusion concerning the variables of integration. Further details concerning the geometry and derivation of Eq. 2 may be found in Chapman and Cowling.[5]

b. Small m_e/m_β Approximation

The system of Eqs. 1 represents, in general, a set of three coupled, nonlinear, six-dimensional equations for the distribution functions $f_\alpha(\bar{r}, \bar{v}_\alpha)$; $\alpha = e, i, n$. This system of equations may be greatly simplified by utilizing the fact that the mass ratio m_e/m_β ($\beta = i, n$) is small and noting that

$$J_{e\beta}(f_e, f_\beta) = J_{e\beta}[f_e, n_\beta \delta(\bar{v}_\beta)] + 0(m_e/m_\beta); \quad \beta = i, n, \tag{3}$$

where n_β is the density of species β, and $\delta(x)$ is the Dirac delta function. Physically, Eq. 3 implies that in the Boltzmann equation for electrons, the heavy particles may, to a good approximation, be regarded as stationary. The result is that the electron equation is decoupled from those for the heavy particles (in velocity space). The rest of this part of the report is devoted to the solution of the Boltzmann equation for electrons.

c. Linearized Boltzmann Equation

In seeking a solution for the electron distribution function $f_e(\bar{r}, \bar{v}_e)$, it is convenient to define a perturbation function, $\phi_e(\bar{r}, \bar{v}_e)$, by means of the equation

$$f_e(\bar{r}, \bar{v}_e) = f_e^0(\bar{r}, \bar{v}_e)[1 + \phi_e(\bar{r}, \bar{v}_e)], \tag{4}$$

where

$$f_e^0(\bar{r}, \bar{v}_e) \equiv n_e(m_e/2\pi k T_e)^{3/2} \exp\left(-m_e v_e^2/2k T_e\right) \tag{5}$$

$$n_e = \int f_e \, d\underline{v}_e; \quad \frac{3}{2} n_e k T_e = \int \left(m_e v_e^2/2\right) f_e \, d v_e \tag{6}$$

Thus, the perturbation ϕ_e must satisfy the conditions

$$\int f_e^0 \phi_e \, d\underline{v}_e = 0; \quad \int v_e^2 f_e^0 \phi_e \, d\underline{v}_e = 0. \tag{7}$$

With these definitions and conditions the linearized Boltzmann equation for electrons is

$$\frac{1}{n_e}f^0_e\left\{\overline{v}_e\cdot\left(\frac{\nabla_r P_e}{P_e}+\frac{eE}{kT_e}\right)+\left(u^2_e-\frac{5}{2}\right)\overline{v}_e\cdot\frac{\nabla_r T_e}{T_e}\right\}=-n_e I_e(\phi_e)-n_i I_{ei}(\phi_e)-n_n I_{en}(\phi_e),\qquad(8)$$

where

$$\overline{u}_e=(m_e/2kT_e)^{1/2}\overline{v}_e\qquad\qquad(9)$$

and the linear integral operators I_e and $I_{e\beta}$ are defined as

$$I_e[F_e(\overline{v}_e)]\equiv\frac{1}{n^2_e}\iiint f^0_e f^0 (F_e+F-F'_e-F')\,g_e b\,db d\epsilon d\underline{v}\qquad\qquad(10)$$

$$I_{e\beta}[F_e(\overline{v}_e)]\equiv\frac{1}{n_e}\iint f^0_e(F_e-F'_e)\,v_e b\,db d\epsilon.\qquad\qquad(11)$$

In definitions (10) and (11), $F_e(\overline{v}_e)$ may be a scalar or a vector function of $\overline{v}_e$, and $f^0=f^0_e(\overline{r},\overline{v})$, $F=F_e(\overline{v})$, $F'_e=F_e(\overline{v}'_e)$, $F'=F_e(\overline{v}')$, and $g_e=|\overline{v}_e-v|$.

d. Solution of the Linearized Boltzmann Equation

The general solution of Eq. 8 is

$$\phi_e=-\frac{1}{n_e}(2kT_e/m_e)^{1/2}\left\{\overline{A}_e(\overline{u}_e)\cdot\left(\frac{\nabla P_e}{P_e}+\frac{e\overline{E}}{kT_e}\right)+\overline{B}_e(\overline{u}_e)\cdot\frac{\nabla T_e}{T_e}\right\}+C_1 m_e+C_2 m_e v^2_e/2,$$

$$(12)$$

where $\overline{A}_e$, $\overline{B}_e$ are vector functions of the electron velocity and C_1, C_2 are arbitrary constants. The first two terms on the right-hand side of Eq. 12 represent the particular solutions for each of the driving terms in Eq. 8, while the last two terms represent the homogeneous solution.

Substitution of Eq. 12 in Eq. 8 yields

$$f^0_e\overline{u}_e=n_e I_e(\overline{A}_e)+n_i I_{ei}(\overline{A}_e)+n_n I_{en}(\overline{A}_e)\qquad\qquad(13)$$

$$\left(u^2_e-\frac{5}{2}\right)f^0_e\overline{u}_e=n_e I_e(\overline{B}_e)+n_i I_{ei}(\overline{B}_e)+n_n I_{en}(\overline{B}_e).\qquad\qquad(14)$$

Since Eqs. 13 and 14 contain only $\overline{u}_e$ as an independent variable, $\overline{A}_e$ and $\overline{B}_e$ must be of the form

$$\overline{A}_e(\overline{u}_e) = A_e(u_e)\overline{u}_e/u_e; \quad \overline{B}_e(\overline{u}_e) = B_e(u_e)\overline{u}_e/u_e, \tag{15}$$

where $A_e(u_e)$ and $B_e(u_e)$ are scalar functions of the magnitude of $\overline{u}_e$.
From Eqs. 7-15 we find that

$$C_1 = C_2 = 0, \tag{16}$$

$$I_{e\beta}(\overline{G}) = \frac{1}{n_e} f_e^o v_e \sigma_{e\beta}(v_e)\overline{G}; \quad \overline{G} = \overline{A}_e, \overline{B}_e, \tag{17}$$

where

$$\sigma_{e\beta}(v_e) \equiv 2\pi \int (1 - \cos x) b \, db \tag{18}$$

is the momentum transfer cross section.

The meaning of Eqs. 12-18 is that the problem of determining the electron distribution function $f_e(\overline{r}, \overline{v}_e)$ is reduced to that of finding two scalar functions $A_e(u_e)$ and $B_e(u_e)$ that are solutions of Eqs. 13 and 14, respectively. These equations can be solved exactly only for a Lorentz plasma (see below). In the general case of a three-component plasma one resorts to approximation techniques which will be discussed below.

e.   Sonine Polynomial Expansions for $A_e(u_e)$ and $B_e(u_e)$

For a general three-component plasma, it is expedient to expand the scalar functions $A_e(u_e)$ and $B_e(u_e)$ into series of Sonine polynomials[7] of order 3/2:

$$A_e(u_e)/u_e = \sum_{n=0}^{\infty} a_n S_n^{3/2}(u_e^2); \quad B_e(u_e)/u_e = \sum_{n=0}^{\infty} b_n S_n^{3/2}(u_e^2), \tag{19}$$

where a_n, b_n are expansion coefficients, and $S_n^{3/2}(x)$ is a Sonine polynomial of order 3/2. Substitution of these expansions in Eqs. 13 and 14, dot-multiplication of the result by $S_m^{3/2}(u_e^2)\overline{u}_e$, and integration over velocity space yields two infinite sets of linear algebraic equations of the form

$$\sum_{n=0}^{\infty} \alpha_{mn} a_n = \beta_m$$

$$m = 0, 1, \ldots \infty, \tag{20}$$

$$\sum_{n=0}^{\infty} \alpha_{mn} b_n = \gamma_m,$$

where

$$a_{mn} \equiv a_{mn}^{ee} + a_{mn}^{ei} + a_{mn}^{en};$$

$$a_{mn}^{ee} \equiv \int S_m^{3/2}\left(u_e^2\right)\overline{u}_e \cdot I_e\left[S_n^{3/2}\left(u_e^2\right)\overline{u}_e\right] d\underline{v}_e;$$

$$a_{mn}^{e\beta} \equiv \frac{n_\beta}{n_e} \int S_m^{3/2}\left(u_e^2\right)\overline{u}_e \cdot I_{e\beta}\left[S_n^{3/2}\left(u_e^2\right)\overline{u}_e\right] d\underline{v}_e; \tag{21}$$

$$\beta_m \equiv \frac{1}{n_e} \int S_m^{3/2}\left(u_e^2\right)u_e^2 f_e^o \, d\underline{v}_e = \frac{3}{2}\delta_{mo},$$

$$\gamma_m \equiv \frac{1}{n_e} \int S_m^{3/2}\left(u_e^2\right)u_e^2\left(u_e^2 - \frac{5}{2}\right)f_e^o \, d\underline{v}_e = -\frac{15}{4}\delta_{m1}.$$

Thus, the problem of solving Eqs. 13 and 14 for the scalar functions $A_e(u_e)$ and $B_e(u_e)$ is reduced to that of solving the two infinite sets of Eqs. 20 for the Sonine expansion coefficients a_n and b_n. Approximate solutions to any desired degree of accuracy may be obtained by truncating the Sonine expansions after N terms and solving the resulting 2N equations. The matrix elements a_{mn}, in principle, may be determined once the collision laws are specified. Specifically, the quantities a_{mn}^{ee} are special cases of a general set of like-particle collision integrals which have been tabulated by Chapman and Cowling,[6] while $a_{mn}^{e\beta}$ can be evaluated by using Eq. 17 for the operator $I_{e\beta}$. It can also be shown that $a_{mn} = a_{nm}$.[6]

f. Electron Current and Heat Flux

The electron current, $\overline{J}_e$, and the electron heat flux, $\overline{q}_e$, are given by

$$\overline{J}_e \equiv e \int \overline{v}_e f_e \, d\underline{v}_e = -\mu_e\left\{\nabla P_e + en_e\overline{E} + k_e^T n_e \nabla k T_e\right\} \tag{22}$$

$$q_e \equiv \int \overline{v}_e \frac{m_e v_e^2}{2} f_e \, d\underline{v}_e = \frac{J_e}{e}\frac{5}{2}kT_e + \frac{\overline{J}_e}{e}k_e^T kT_e - \mathcal{K}_e\nabla T_e, \tag{23}$$

where

$$\mu_e \equiv \frac{2e}{3n_e^2 m_e} \int f_e^o u_e A_e(u_e) \, d\underline{v}_e \qquad \text{(electron mobility)} \tag{24}$$

$$k_e^T \equiv \frac{2e}{3n_e^2 m_e \mu_e} \int f_e^o u_e B_e(u_e) \, d\underline{v}_e \qquad \text{(thermal diffusion ratio)} \tag{25}$$

$$\mathcal{K}_e \equiv \frac{\mu_e n_e k^2 T_e}{e} \left\{ \frac{2e}{3n_e^2 m_e \mu_e} \int f_e^o u_e^3 B_e(u_e)\, d\underline{v}_e - k_e^T \left(\frac{5}{2} + k_e^T\right) \right\} \qquad \text{(thermal conductivity)}$$

(26)

The last term in (22) is the "thermal diffusion" term of Chapman and Cowling and other authors. The first term on the right-hand side of (23) represents enthalpy transport, the second term is the "diffusion thermo-effect," an energy transport mechanism related to the thermal diffusion mechanism in Eq. 22, and the third term accounts for thermal conduction. The definitions of the transport coefficients μ_e, k_e^T, and $\mathcal{K}_e$ are consistent with the usual definitions of these quantities.

In terms of the Sonine expansion coefficients Eqs. 24-26 become

$$\mu_e = \frac{e a_o}{n_e m_e}; \qquad k_e^T = \frac{b_o}{a_o}; \qquad \mathcal{K}_e = -\frac{5}{2} \frac{k^2 T_e}{m_e} \left(b_1 + \frac{2}{5} k_e^T b_o\right).$$

(27)

3. Applications of the Perturbation Method

a. Lorentz Plasmas

For a Lorentz plasma Eqs. 13 and 14 may be solved exactly. Hence the Lorentz plasma provides a convenient reference case for comparisons. Specifically, for a Lorentz plasma consisting of electrons and heavy particles of species β only, the perturbation function and transport coefficients are

$$\phi_e = \lambda_{e\beta} \left\{ \frac{\nabla P_e}{P_e} + \frac{e\overline{E}}{kT_e} + \left(u_e^2 - \frac{5}{2}\right) \frac{\nabla T_e}{T_e} \right\} \cdot \frac{\overline{u}_e}{u_e}$$

(28)

$$\mu_e = \frac{e}{P_e} \int \frac{v_e^2}{3\nu_{e\beta}} f_e^o\, dv_e$$

$$k_e^T = \frac{e}{\mu_e P_e} \int \frac{v_e^2}{3\nu_{e\beta}} \left(\frac{m_e v_e^2}{2kT_e} - \frac{5}{2}\right) f_e^o\, d\underline{v}_e$$

(29)

$$\mathcal{K}_e = \frac{\mu_e k P_e}{e} \left\{ \frac{e}{\mu_e P_e} \int \frac{v_e^2}{3\nu_{e\beta}} \left(\frac{m_e v_e^2}{2kT_e}\right)^2 f_e^o\, d\underline{v}_e - \left(\frac{5}{2} + k_e^T\right)^2 \right\},$$

where $\lambda_{e\beta}(v_e) = 1/n_\beta \sigma_{e\beta}(v_e)$ is the electron mean-free path, and $\nu_{e\beta}(v_e) = v_e/\lambda_{e\beta}(v_e)$ is is the e-β collision frequency. Note that the perturbation ϕ_e is small (compared with unity) when the fractional changes in the electron pressure and temperature and in the plasma potential are small over one mean-free path.

Table XI-2. Special values of the transport coefficients for Lorentz plasmas.

Heavy-Particle Species	Neutrals	Neutrals	Ions
Collision Law	ν_{en} = const	λ_{en} = const	Coulomb
μ_e	$\dfrac{e}{m_e \nu_{en}}$	$\dfrac{e\lambda_{en}v_a}{3kT_e}$	$128\left(\dfrac{\pi}{2}\right)^{1/2} \dfrac{\epsilon_0^2(kT_e)^{3/2}}{n_i e^3 m_e^{1/2} \ln \Lambda}$
k_e^T	0	$-1/2$	$+3/2$
$\mathcal{K}_e$	$\dfrac{5}{2}\dfrac{n_e k^2 T_e}{m_e \nu_{en}}$	$\dfrac{2}{3}n_e k\lambda_{en}v_a$	$512\left(\dfrac{\pi}{2}\right)^{1/2} \dfrac{n_e k\epsilon_0^2(kT_e)^{5/2}}{n_i e^4 m_e^{1/2} \ln \Lambda}$

Note: v_a = electron mean speed = $(8kT_e/\pi m_e)^{1/2}$; $\ln \Lambda$ = coulomb logarithm.

Shown in Table XI-2 are values of the transport coefficients obtained from Eqs. 29 for three Lorentz plasmas of special interest: (i) the heavy particles are neutrals and ν_{en} = constant; (ii) the heavy particles are neutrals and λ_{en} = constant; (iii) the heavy particles are ions. A Lorentz plasma of the third type is a purely hypothetical case, since the neglect of e-e collisions as compared with e-i collisions requires that $n_e \ll n_i$, a condition that is never achieved in practice. Note that the values in Table XI-2 are in agreement with similar results derived by other techniques.

The transport coefficients for a Lorentz plasma can also be computed by means of the Sonine polynomial expansion technique. For example, for a Lorentz plasma in which the heavy particles are neutrals and λ_{en} = constant, these coefficients for N = 1, 2,

Table XI-3. Transport coefficients for Lorentz plasma with constant λ_{en}.

Terms in Sonine Expansion	$\mu_e\big/(\mu_e)_{exact}$	$k_e^T\big/(k_e^T)_{exact}$	$\mathcal{K}_e\big/(\mathcal{K}_e)_{exact}$
N = 1	0.88	0	0
N = 2	0.95	0.77	0.85
N = 3	0.98	0.90	0.93

and 3 can be readily evaluated and compared with the exact values given in Table XI-2. This comparison reveals (Table XI-3) that retention of three terms in each of the Sonine expansions is sufficient to yield all three transport coefficients correctly within 10 per cent.

b. Three-Component Plasmas with Constant v_{en}.

Another illustration of the application of the Sonine polynomial expansion technique is the computation of the transport coefficients for a three-component plasma with v_{en} = constant and $n_e \approx n_i$. If three terms are again retained in each of the expansions, the matrix elements a_{mn} can be readily evaluated, and they lead to the following set of equations:

$$\begin{bmatrix} 1 + 1.50 v_{en}/v_{ei} & 1.50 & 1.87 \\ 1.50 & 4.66 + 3.75 v_{en}/v_{ei} & 5.37 \\ 1.87 & 5.37 & 10.7 + 13.1 v_{en}/v_{ei} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1.50 n_e/v_{ei} \\ 0 \\ 0 \end{bmatrix}$$

(30)

$$\begin{bmatrix} 1 + 1.50 v_{en}/v_{ei} & 1.50 & 1.87 \\ 1.50 & 4.66 + 3.75 v_{en}/v_{ei} & 5.37 \\ 1.87 & 5.37 & 10.7 + 13.1 v_{en}/v_{ei} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0 \\ -3.75 n_e/v_{ei} \\ 0 \end{bmatrix}$$

(31)

Here, $v_{ei} \equiv \dfrac{e^4 \ln \Lambda v_a n_i}{4\pi \epsilon_0^2 (2kT_e)^2}$ is the effective electron-ion collision frequency.

Solution of Eqs. 30-31 yields a_n, b_n which in turn can be used in Eqs. 27 to calculate the transport coefficients. The results of the calculation are given in Table XI-4.

The second column of Table XI-4 gives the values of the coefficients in the Lorentz limit $v_{ei} \ll v_{en}$, denoted by the superscript "o"; they are identical to the corresponding values in Table XI-2. The reason is that the Sonine expansion technique with $N \geq 2$ yields an exact solution for the Lorentz plasma with constant v_{en}.

The third column of Table XI-4 gives the values of the electron transport coefficients in the fully ionized limit $v_{ei} \gg v_{en}$. They are denoted by the superscript "∞", and are in excellent agreement with results reported by Chapman and Cowling[3] and by Samokhin.[8] By comparing the values for μ_e^∞, $\left(k T_e\right)^\infty$, and $\mathcal{K}_e^\infty$ with the corresponding values in Table XI-2 for a hypothetical e-i Lorentz plasma, the importance of electron-electron collisions is deduced. The effect of e-e collisions is to reduce each of the transport coefficients (in the fully ionized limit) approximately from one-fourth to one-half the value obtained in the absence of e-e collisions.

Table XI-4.   Transport coefficients in a three-component plasma.

	Lorentz Limit $(\nu_{ei} \ll \nu_{en})$	Fully Ionized Limit $(\nu_{ei} \gg \nu_{en})$	General Expression
μ_e	$\mu_e^o = \dfrac{e}{m_e \nu_{en}}$	$\mu_e^\infty = 0.57 \times 128 \left(\dfrac{\pi}{2}\right)^{1/2} \dfrac{\epsilon_o^2 (kT_e)^{3/2}}{e^3 n_i m_e^{1/2} \ln \Lambda}$	$\mu_e = \dfrac{\mu_e^o \mu_e^\infty}{\mu_e^o + \mu_e^\infty} h_\mu(\mu)$
k_e^T	$\left(k_e^T\right)^o = 0$	$\left(k_e^T\right)^\infty = 0.71$	$k_e^T = \left(k_e^T\right)^\infty h_k(\mu)$
$\mathcal{K}_e$	$\mathcal{K}_e^o = \dfrac{5}{2} \dfrac{n_e k^2 T_e}{m_e \nu_{en}}$	$\mathcal{K}_e^\infty = 0.23 \times 512 \left(\dfrac{\pi}{2}\right)^{1/2} \dfrac{k \epsilon_o^2 (kT_e)^{5/2}}{e^4 m_e^{1/2} \ln \Lambda}$	$\mathcal{K}_e = \dfrac{\mathcal{K}_e^o \mathcal{K}_e^\infty}{\mathcal{K}_e^o + \mathcal{K}_e^\infty} h_{\mathcal{K}}(\mu)$

The fourth column of Table XI-4 lists the values of the coefficients for values of ν_{ei}/ν_{en} lying between the Lorentz and fully ionized limits.   The functions $h_\mu(\mu)$, $h_k(\mu)$, and $h_{\mathcal{K}}(\mu)$ are given by

$$h_\mu(\mu) = \frac{1.00 + 7.00\mu + 9.67\mu^2 + 3.67\mu^3}{1.00 + 7.93\mu + 10.9\mu^2 + 3.67\mu^3}$$

$$h_k(\mu) = \frac{1.00\mu + 0.89\mu^2}{0.24 + 1.46\mu + 0.89\mu^2} \tag{32}$$

$$h_{\mathcal{K}}(\mu) = \frac{1.00 + 11.9\mu + 45.1\mu^2 + 73.2\mu^3 + 52.0\mu^4 + 13.5\mu^5}{1.00 + 13.9\mu + 62.0\mu^2 + 95.0\mu^3 + 65.9\mu^4 + 13.5\mu^5},$$

where

$$\mu \equiv \frac{\mu_e^o}{\mu_e^\infty} = 0.34 \frac{\nu_{ei}}{\nu_{en}}.$$

These functions are plotted in Fig. XI-42.

From Fig. XI-42 it is apparent that the function $h_\mu(\mu)$ is nearly unity for all values of μ $(0 \leqslant \mu \leqslant \infty)$.   This implies that, to a good approximation, the contributions to the electron mobility from electron-neutral and from electron-charged particle collisions may be added in parallel.   Similar conclusions apply to the contributions to the thermal conductivity of the electrons.
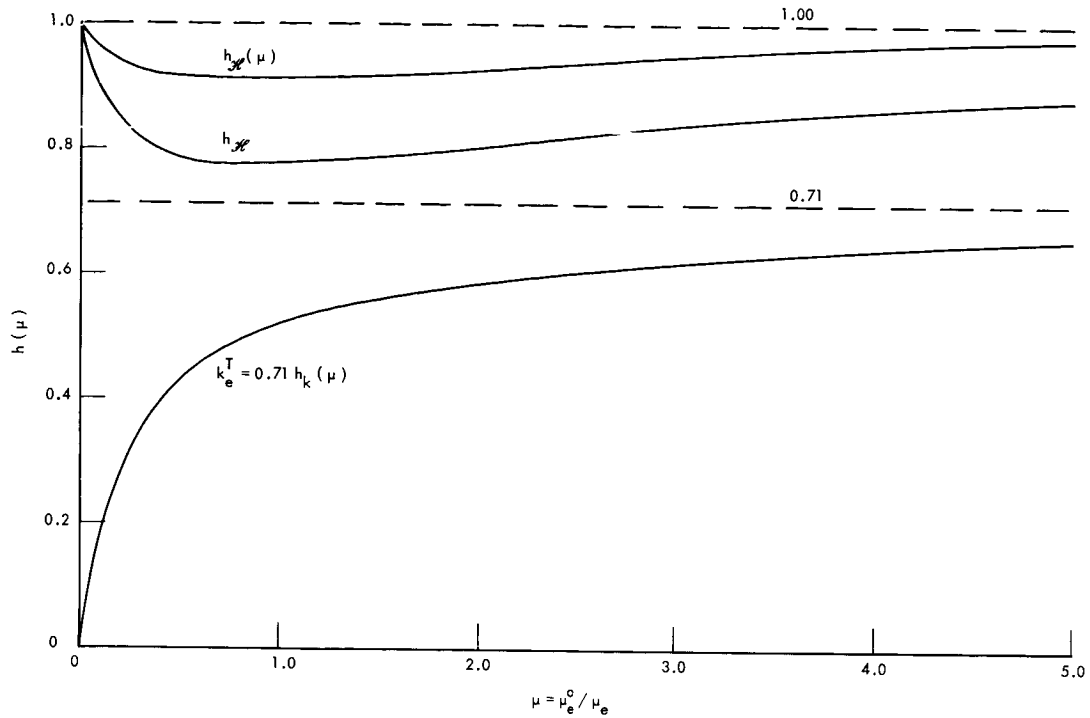
Fig. XI-42. Plots of h(μ) functions against μ.

4. Conclusions

The modified Chapman-Cowling approach presented in this report provides a useful tool for the analysis of three-component plasmas. The method yields both a quantitative description of the electron-particle and energy-transport mechanisms and an analytical expression for the electron distribution function.

D. R. Wilkins, E. P. Gyftopoulos

References

1. H. Grad, Communs. Pure Appl. Math. 2, 324 (1949).

2. J. O. Hirshfelder, C. F. Curtiss, and R. B. Bird, Molecular Theory of Gases and Liquids (John Wiley and Sons, Inc., New York, 1954).

3. I. P. Stachanov and R. S. Stepanov, Soviet Phys. – Tech. Phys. 9, 315 (1964).

4. S. Chapman and T. G. Cowling, The Mathematical Theory of Nonuniform Gases (Cambridge University Press, London, 1961).

5. Ibid, p. 65.

6. Ibid, pp. 85-88; 161.

7. Ibid, p. 124.

8. M. V. Samokhin, Soviet Phys. – Tech. Phys. 8, 498 (1963).

Q. PROBLEMS IN THE THEORY OF OPTIMAL CONTROL OF NONLINEAR SYSTEMS

The existence of solutions to certain optimal control problems has been studied for a multi-input, multi-output nonlinear system that is realizable, and is representable in terms of an input-output functional series of the form

$$\underline{y}(t) = \sum_{j=0}^{\infty} \sum_{i=1}^{i_{j,r}} \int_{0}^{t-t_o} \cdots \int \underline{h}_{j,i}(\sigma_1, \sigma_2, \ldots, \sigma_j) \, C_i(\underline{u}; \sigma_1, \sigma_2, \ldots, \sigma_j) \, d\sigma_1 d\sigma_2 \cdots d\sigma_j \qquad (1)$$

which is an extension of the system representation of Chesler[1] to multi-output case. The optimal control problem in this formulation is to find the input vector $\underline{u}(t) \in U$, $t_o \leq t \leq t_1$, which is such that it satisfies the system and the boundary conditions

$$\underline{\eta} = \sum_{j=0}^{\infty} \sum_{i=1}^{i_{j,r}} \int_{0}^{t_1-t_o} \cdots \int \underline{h}_{j,i}(\sigma_1, \sigma_2, \ldots, \sigma_j) \, C_i(\underline{u}; \sigma_1, \sigma_2, \ldots, \sigma_j) \, d\sigma_1 d\sigma_2 \cdots d\sigma_j \qquad (2)$$

and minimizes the cost functional

$$\mathscr{J}(\underline{u}) = \int_{t_o}^{t_1} \mathscr{L} \left[\underline{u}(t), \sum_{j=0}^{\infty} \sum_{i=0}^{i_{j,r}} \int_{0}^{t-t_o} \cdots \int \underline{h}_{j,i}(\sigma_1, \sigma_2, \ldots, \sigma_j) \right.$$

$$\left. \cdot \, C_i(\underline{u}; \sigma_1, \sigma_2, \ldots, \sigma_j) \, d\sigma_1 d\sigma_2 \cdots d\sigma_j \right] dt, \qquad (3)$$

where U is the specified subset of the input space, the n-vector $\underline{\eta}$ is the desired terminal value of the output, the n-vector $\underline{h}_{j,i}(\sigma_1, \sigma_2, \ldots, \sigma_j)$ is the ith component of the jth-order kernel of the nonlinear system, and $C_i(\underline{u}; \sigma_1, \sigma_2, \ldots, \sigma_j)$ is the ith possible j-tuple product of the components of the input vector $\underline{u}(t)$, each factor in the product having σ_1, σ_2, ..., or σ_j as its argument.

We find the set V of the inputs that satisfy Eq. 2, and determine the set $W = V \cap U$. The element $\underline{u}^* \in W$ that minimizes the cost functional, Eq. 3, will then be the optimal input that is sought. If the set V is empty or the sets V and U are disjoint, that is, if the set W is empty, we conclude that no input exists that can perform the prescribed task. If W is nonempty and compact, we ascertain the existence of the optimal input. If W contains only a single element or the element $\underline{u}^* \in W$ that minimizes Eq. 3 is unique, we say that the optimal control problem has a unique solution.

In obtaining the set V, the potentialness of the system operator of Eq. 2 is necessary. The conditions that the kernels $\underline{h}_{j,i}$ must satisfy in order for the system operator to be potential, as well as the potential of the operator, have been found. It has also been found that the set $\bar{V}$, if it is not empty, is compact.

The application of the theory to systems that can be represented with only a few terms of the series is now being studied. For these systems, the possibility of developing a computational technique of obtaining the optimal input with respect to some practical cost functionals will be investigated by using the present formulation.

S. H. Kyong

References

1. D. A. Chesler, Nonlinear Systems with Gaussian Inputs, Technical Report 366, Research Laboratory of Electronics, M. I. T., Cambridge, Mass., February 15, 1960.

2. M. M. Vainberg, Variational Methods for the Study of Nonlinear Operators, (Holden-Day, Inc., San Francisco, 1964).

XII. PLASMA MAGNETOHYDRODYNAMICS AND ENERGY CONVERSION[*]

A. EXPERIMENTS WITH A LIQUID-METAL MAGNETOHYDRODYNAMIC WAVEGUIDE

There have been several attempts in the last few years to show the existence of Alfvén waves in a liquid-metal waveguide. Gothard[1,2] designed an experimental waveguide and, using it as a resonator, obtained preliminary data. His equipment was used by Jackson and Carson[3] to obtain more conclusive and extensive data. These investigators were able to obtain field-dependent resonances in the waveguide at frequencies corresponding to those of TM Alfvén wave modes. They were not able, however, to see anything conclusive with a coil probe which was located in the waveguide cavity.

In this investigation, the objective was to display the waves in the most direct manner so that it could be used in a film, entitled "Magnetohydrodynamics," which is now being made for the Fluid Mechanics Film Series produced by Educational Services Incorporated. To reduce attenuation, the walls of the Gothard waveguide were lined with copper and the ends were insulated to obtain boundary conditions that correspond closely to the theoretical model used by Reid.[4] The excitation system was also changed from a disk exciter, which produces TM modes, to a copper cylindrical center conductor, which produces TEM modes. Thus the waveguide was essentially a coaxial line filled with a liquid metal. Sodium-potassium alloy (NaK) was again used as the working fluid. The excitation current was changed to a square wave so that the progress of the wave front past two stationary probe coils could be observed.

The complete waveguide assembly is shown in Figs. XII-1 through XII-4; it can be seen that the waveguide is a copper can fitting into the original stainless-steel structure developed by Gothard. The center post is made of solid copper with an outside diameter of 2.5 cm, and the length of the waveguide is 15 cm. This is 3 cm smaller than the total length of the magnet gap, and allows clearance for the input cables to be attached to the
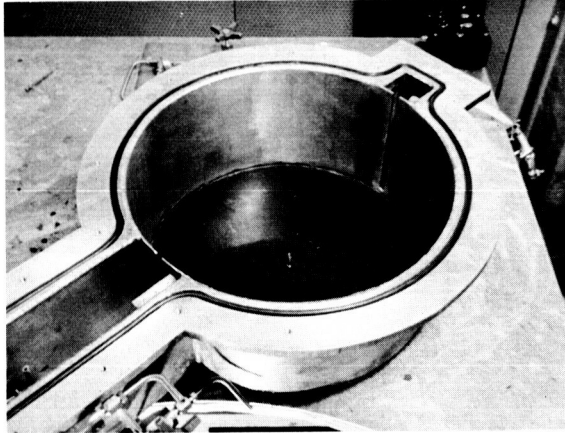
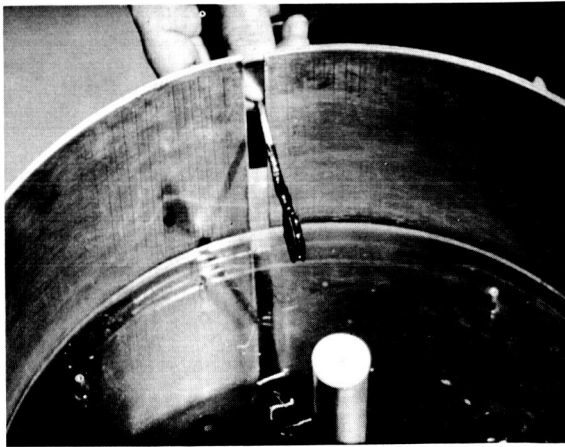Fig. XII-1. Gothard waveguide showing insulated lower end.

Fig. XII-2. Copper insert with center post.

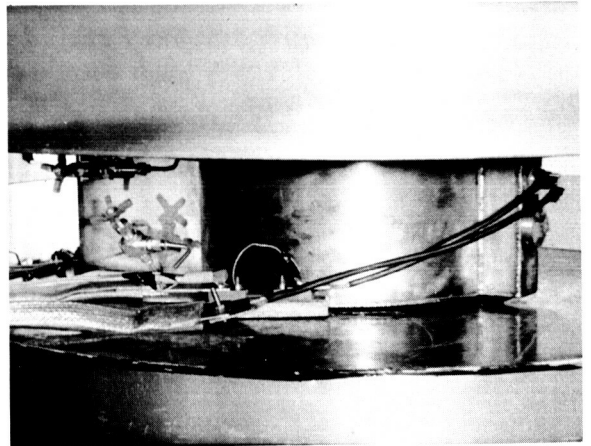Fig. XII-3. Search coil position in the waveguide cavity.

Fig. XII-4. Waveguide positioned in magnetic field.

center of the guide to obtain current distribution independently of the ϕ-coordinate. The field probe consisted of a solenoid, 3/8 inch in diameter, with 600 turns of No. 37 wire laid in epoxy. The output of the coils was isolated from the rest of the system by transformers to eliminate the effects of ground currents from the measuring circuits. The transformer output was then amplified approximately 40 db, and the final output was displayed on a Tektronix oscilloscope.

The applied magnetic field was 0.8 weber/meter^2, and the exciting current was a 10-cps square wave with an amplitude of 100 amps. The fluid properties of the NaK alloy used in the experiment were: density $\rho = 0.85 \times 10^3$ kg/m^3; electrical conductivity $\sigma = 2.4 \times 10^6$ mhos/meter. The Alfvén wave velocity v_a is given by $v_a = B_0/\sqrt{\mu_0 \rho}$, which, under the experimental conditions used, was calculated to be $v_a = 24.5$ meters/second. Thus the wave should transverse the length of the waveguide in

LOWER PROBE　　　　UPPER PROBE
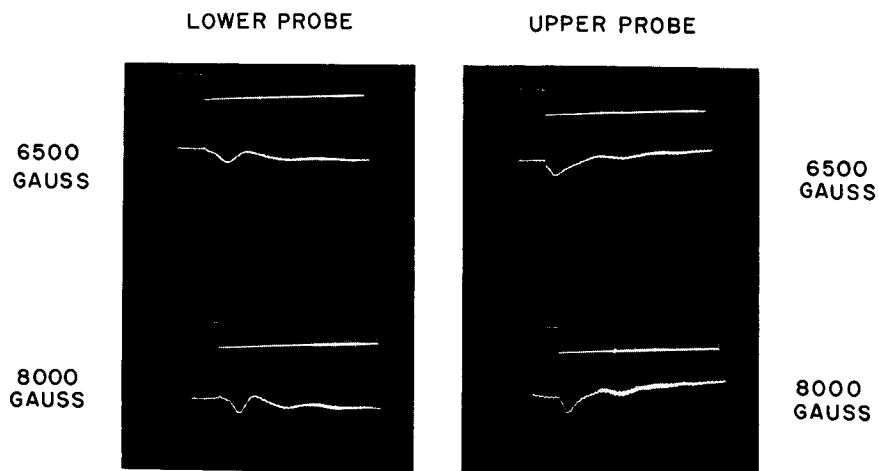

6500
GAUSS

6500
GAUSS

8000
GAUSS

8000
GAUSS

Fig. XII-5. Drive current and probe waveforms. Upper trace: exciting current. Lower trace: probe voltage waveform. (Time scale, 5 msec/cm.)

6.2 msec. Reproductions of the typical oscilloscope traces, obtained by using a Polaroid camera, are shown in Fig. XII-5. Reflections are clearly visible and serve to confirm the fact that the observed phenomenon has the character of wave motion. The arrival times of the initial wavefront at the first (upper) and second (lower) probes were 2.34 msec and 4.90 msec, respectively. The measured velocity was estimated to be 23.0 m/sec, and the upper and lower probes were calculated to be 5.7 cm and 11.9 cm, respectively, from the top of the waveguide. Good agreement was obtained with the predicted behavior of a TEM Alfvén mode in a cylindrical waveguide, and an error of ~6 per cent arises largely from the difficulty of locating the exact arrival time in the presence of a substantial amount of wave dispersion.

W. D. Jackson, B. D. Wessler, G. B. Kliman

References

1. N. Gothard, S. M. Thesis, Department of Electrical Engineering, M. I. T., 1962.

2. N. Gothard, Excitation of hydromagnetic waves in a highly conducting liquid, Phys. Fluids 7, 1784 (1964).

3. W. D. Jackson and J. F. Carson, Quarterly Progress Report No. 72, Research Laboratory of Electronics, M. I. T., January 15, 1964, p. 149.

4. M. H. Reid, S. B. Thesis, Department of Electrical Engineering, M. I. T., 1962.

B. HALL PARAMETER — CONDUCTIVITY INSTABILITIES IN MAGNETOGASDYNAMIC FLOW

McCune[1] has analyzed instabilities in a slightly ionized plasma, which are due to fluctuations in the local conductivity and Hall parameter. Essentially, his work expanded the analysis of Hall parameter instabilities previously conducted by Velikhov.[2] McCune's analysis differed from Velikhov's with respect to both model and approach: McCune used a dispersion relation approach, while Velikhov used an energy approach. It appears that stability criteria are most readily obtained for McCune's model by using Velikhov's approach.

Consider a plasma with a low magnetic Reynolds number, slight ionization (less than 0.1 per cent), in the presence of a strong magnetic field so that the Hall parameter is greater than unity. The pertinent linearized first-order electromagnetic, fluid-mechanic, and constitutent equations are listed below. All terms, except physical constants that have zero subscripts, indicate zero-order terms. The wave disturbance is in a plane perpendicular to the magnetic field, parallel or antiparallel to the current, one-dimensional, and proportional to $\exp[i(\omega t - \bar{k} \cdot \bar{r})]$.

1. $\dfrac{\partial \rho}{\partial t} + \rho_0 \nabla \cdot \bar{v} = 0$ $\qquad \Omega_0 = \dfrac{w}{\nu_0}$

2. $\nabla \cdot \bar{j} = 0$ $\qquad w = $ cyclotron frequency

3. $\nabla \times \bar{e} = 0$ $\qquad \nu_0 = \dfrac{1}{\tau} = $ collision frequency

4. $\rho_0 \dfrac{\partial v}{\partial t} + \nabla p = \bar{j} \times \bar{B}_0$ $\qquad \Omega_0 = $ Hall parameter

5. $\bar{j} + \bar{j} \times \bar{\Omega}_0 + \bar{J}_0 \times \bar{\Omega} = \sigma_0(\bar{e} + \bar{v} \times \bar{B}_0) + \sigma \bar{E}_0$

If we combine the x component of the last equation with the remaining four equations we find that

$$jx = -J_o \bar{\Omega}_o \left(\frac{\Omega}{\Omega_o}\right) + \sigma_o E_{ox} \left(\frac{\sigma}{\sigma_o}\right) + \sigma_o B_o v.$$

The perturbation current is composed of three parts: perturbation currents caused by (a) local fluctuations in the Hall parameter, (b) fluctuations in the conductivity, and (c) the electromotive force. This total perturbation current, together with the uniform magnetic field, will produce a force that will cause instability if the force and velocity are in the same direction. For a slightly ionized gas the conductivity will increase as the temperature to the power a (where a is ~13). Also,

$$\nu_o = \left(\frac{8kT_o}{\pi M_e}\right)^{1/2} N_o Q_N,$$

where Q_N is the collision cross section between electrons and neutrals, and N_o is the neutral density. Therefore,

$$\frac{\sigma}{\sigma_o} = a \frac{T}{T_o}$$

and

$$\frac{\Omega}{\Omega_o} = -\frac{1}{2} \frac{T}{T_o} - \frac{N}{N_o}.$$

It can be shown that for small temperature variations the conductivity also depends upon density variations so that

$$\frac{\sigma}{\sigma_o} = a \frac{T}{T_o} - \frac{1}{2} \frac{\rho}{\rho_o}.$$

The density dependence can be neglected whenever $a \approx 13$ and $\gamma = \dfrac{C_p}{C_v}$ is greater than approximately 1.2.

If we use this relationship and assume the equation of state for an ideal gas, we find that

$$-j_x B_o \left\{1 + i\frac{1}{2}\frac{J_o B_o \Omega_o}{p_o k} + i\frac{aE_{ox} B_o \sigma_o}{p_o k}\right\} = -\frac{J_o B_o \Omega_o}{2}\left[\gamma\left(\frac{v_{ph}}{a_o}\right)^2 + 1\right]\frac{k}{\omega}v - \sigma_o B_o^2 v$$

$$-aE_{ox} B_o \sigma_o \left[\gamma\left(\frac{v_{ph}}{a_o}\right)^2 - 1\right]\frac{k}{\omega}v.$$

If we consider short wavelengths such as

$$\frac{a J_o B_o \Omega_o}{p_o k} \ll 1,$$

or equivalently

$$\lambda \ll \frac{2\pi p_o}{J_o B_o \Omega_o} = L_i,$$

where L_i is an interaction length, the resulting energy equation is

$$j_x v^* B_o = |v|^2 \left\{ -\sigma_o B_o^2 - \frac{J_o B_o \Omega_o}{2} \left[\gamma \left(\frac{v_{ph}}{a_o}\right)^2 + 1 \right] \frac{k}{\omega} \right.$$

$$\left. - J_o B_o \Omega_o a \left[\gamma \left(\frac{v_{ph}}{a_o}\right)^2 - 1 \right] \frac{k}{\omega} \right\}.$$

The terms on the right-hand side are due to the electromotive force, Hall parameter fluctuation, and conductivity fluctuation perturbation currents, respectively. For short wavelengths such that

$$\lambda \ll 2\pi \frac{(\rho_o C_o T_o) \, v_{ph}}{\left(J_o^2 / \sigma_o\right) \Omega_o}$$

the phase velocity is approximately equal to the speed of sound. For $\gamma = 1$ and for stability the current density must satisfy

$$J_o < \frac{\sigma_o v_{ph} B_o}{\Omega_o} = J_{CR}.$$

This is Velikhov's criterion; it is recovered here because we have assumed that the conductivity is a function only of temperature.

If we assume that $\frac{1}{2}(\gamma+1)$ is almost one order of magnitude less than $a(\gamma-1)$, for stability, we have

$$J_o < \frac{J_{cr}}{a(\gamma-1)}.$$

In this case the conductivity fluctuation instability dominates. In fact, we see that this will be the case whenever

$$\gamma > \frac{2a + 1}{2a - 1}.$$

Recently, McCune[3] arrived at similar conclusions by analyzing growth rates governed by an appropriate dispersion relation.

We conclude that for practical cases of interest the conductivity fluctuation instability will be present and will dominate whenever a critical current density is exceeded. It should be emphasized that this conductivity instability would not exist without the presence of the Hall effect.

K. R. Edwards

References

1. J. E. McCune, Wave-Growth and Instability in Partially Ionized Gases, Report AMP 136, Avco-Everett Research Laboratory, 1964.

2. E. P. Velikhov, Hall Instability of Current Carrying Slightly Ionized Gases, Symposium on Magnetoplasmadynamic Electric Power Generation, Kings College, University of Durham, Newcastle-upon-Tyne, England, 1962.

3. J. E. McCune, Linear Theory of an MHD Oscillator, Report 198, Avco-Everett Research Laboratory, December 1964.

C. NONLINEAR EFFECTS OF FLUCTUATIONS ON MAGNETOHYDRODYNAMIC PERFORMANCE

Recently, attention[1-4] has been given to the possibility of the appearance of certain types of wave amplification mechanisms ("instabilities") in slightly ionized plasmas, and to the question of their importance in magnetohydrodynamic applications. Among the instabilities investigated thus far, in particular, two stand out as potentially important for MHD machines operating at moderate to high Hall parameters. One of these, the wave instability first noted by Velikhov,[2] and further discussed by the present author,[3] involves coupling between acoustic wave modes and the Hall effect, and can have a sufficiently high growth rate to be significant in contemplated devices.[3,5] This acoustic instability should be observed most strongly (although not exclusively) whenever the electrons are moderately well-coupled energetically to the heavier species of the gas.[3] In contrast, the second kind of instability, first noted by Kerrebrock,[4] is potentially important when the electrons are only weakly coupled to the heavy neutrals in a slightly ionized gas. In the last case, wave motion (involving only the electrons in the limit of very weak coupling) becomes possible through the competition between the electric field and gradient-induced electron diffusion, and amplification of this motion can occur in the presence of a magnetic field.

It has recently been demonstrated[5] that the Velikhov instability is potentially important in large, combustion-driven MHD devices operating with high Hall coefficients,

especially with certain geometries, and at particular acoustic resonant frequencies. Similarly, the Kerrebrock instability is likely to play an especially important role if significant nonequilibrium ionization, for example, is seeded, noble-gas plasmas, is realized. It should be emphasized, however, that the acoustic (Velikhov) instability can also be important in slightly ionized plasmas of this last kind because, even with weak energy coupling between the electrons and heavy species, there is still a direct dependence of the Hall parameter on the gas density.[2,3] This means that acoustic fluctuations — however induced — may have important effects on the DC performance of any MHD device that is intended for use with high Hall parameters.

It should be recognized that experience has shown[6] that the moderate and large-scale combustion-driven MHD generators (of the Faraday type and having linear geometry) that have been run actually operate relatively smoothly. The turbulence level in such devices does not appear to be higher than that expected from normal gasdynamic effects. This is consistent with the author's calculations,[5] which have shown in effect that for linear MHD Faraday-type generators, the generator "system" (plasmas, plus boundaries and external circuitry) is stable at all frequencies to the Velikhov mode, below a certain value of the product $\Omega_o M_o h$, where Ω_o is the DC value of the Hall coefficient, M_o is the DC Mach number, and h is the duct height. (A more precise statement of this condition has been given elsewhere.[5]) Until the present time, MHD generators have either been sufficiently small or have been operated at sufficiently low values of $\Omega_o M_o$ to avoid instability even at the resonant frequencies previously discussed.[5]

This result, however, is strongly dependent on the boundary conditions inherent in Faraday-type generators of duct geometry. Such boundary conditions lead in general to quick reflection and damping of the dominant unstable modes (except at resonance[5]).

Such over-all stability of the "system" cannot be expected with certain Hall generator or accelerator geometries, for which the cathode does not present a simple reflecting surface. Indeed, Klepeis and Rosa[7] have observed strong fluctuations at acoustic frequencies in an MHD disc Hall generator.

We are led to conclude that instabilities of the Velikhov type may be important for any sufficiently large MHD device at particular (resonant) frequencies; and for certain geometries (especially Hall generators) they may appear over a much broader frequency spectrum. Less is known about the Kerrebrock instability, but it seems likely that similar conclusions may be reached.

It is important to investigate the possible effect of fluctuations of this kind upon the DC (time-averaged) behavior of MHD devices. This investigation will also be decisive in establishing the maximum AC power level of MHD generator oscillators of the type already discussed.[5]

1. Recent Results

When an MHD device is subject to AC fluctuations, the steady-state amplitude of the "unstable" modes is determined, in conjunction with the external circuitry, by certain internal nonlinear effects which limit the wave amplitudes. These effects are present, as we shall show, because the DC (time-averaged) fields are affected by nonlinear wave interaction, and the wave amplitudes and other wave characteristics, in turn, are functions of the DC fields. For example, the AC impedance of an MHD oscillator,[5] through such effects, becomes a function of the AC amplitudes, and its AC power output can be determined for any given external AC load.

In a recent paper[8] the present author has presented an approximate theory, analogous to the elementary theory of turbulence in ordinary gasdynamics, as a means of relating the DC electric fields and DC currents to the magntiude of the AC fluctuations. More precisely, it is shown that the main power-carrying DC current and the DC "Hall field" are both reduced by correlations of the AC fluctuations. This results physically because the time average of the wave interactions induces local nonuniformities of the type discussed by Rosa,[9] and allows local DC "Hall currents" to flow internally in the MHD device.

The main results can be summarized by the following equations:

$$E'_{xoo} \cong \Omega_{oo} E'_{yoo} \left\{ 1 - \frac{\langle\!\langle \sigma E'_x \rangle\!\rangle_s - \Omega_{oo} \langle\!\langle \sigma E'_y \rangle\!\rangle_s}{\Omega_{oo} \sigma_{oo} E'_{yoo}} + \frac{\langle\!\langle \sigma \Omega \rangle\!\rangle_s}{\sigma_{oo} \Omega_{oo}} - \frac{2 \langle\!\langle \Omega^2 \rangle\!\rangle_s}{1 + \Omega^2_{oo}} \right.$$

$$\left. + \frac{1}{1 + \Omega^2_{oo}} \left[\frac{\left(1 - \Omega^2_{oo}\right) \langle\!\langle \Omega E'_y \rangle\!\rangle_s + 2 \Omega_{oo} \langle\!\langle \Omega E'_x \rangle\!\rangle_s}{\Omega_{oo} E'_{yoo}} \right] \right\}$$

and

$$J_{yoo} \cong \sigma_{oo} E'_{yoo} \left\{ 1 + \frac{\langle\!\langle \sigma E'_y \rangle\!\rangle_s}{\sigma_{oo} E'_{yoo}} - \frac{\Omega^2_{oo}}{1 + \Omega^2_{oo}} \frac{\langle\!\langle \Omega E'_y \rangle\!\rangle_s}{\Omega_{oo} E'_{yoo}} + \frac{\Omega_{oo}}{1 + \Omega^2_{oo}} \frac{\langle\!\langle \Omega E'_x \rangle\!\rangle_s}{\Omega_{oo} E'_{yoo}} - \frac{\langle\!\langle \Omega^2 \rangle\!\rangle_s}{1 + \Omega^2_{oo}} \right\}.$$

In these equations quantities without subscripts represent fluctuations, the angular brackets represent a time average, and $\langle \ \rangle_s$ represents a spatial average. Quantities with double subscript oo are <u>spatial</u> averages of the DC (time-averaged) quantities. For example, if $\hat{j}_y$ is the current density in the y-direction, then $\hat{j}_y = J_{yo} + j_y$, and $\langle \hat{j}_y \rangle = J_{yo}$, $\langle j_y \rangle = 0$, and $\langle J_{yo} \rangle_s = J_{yoo}$. $\hat{\Omega} \equiv \Omega_o + \Omega$ is the local value of $\omega\tau$ in the usual notation, and $\Omega_o \equiv \langle \hat{\Omega} \rangle$. $\underline{E}' \equiv \underline{E}'_o + \underline{E}'$ is the electric field in fluid cooridnates.

The coordinate system is so chosen that the y axis lies in the direction of the "desired" current, so that $\langle J_{xo} \rangle_s \equiv J_{xoo} = 0$, which determines E'_{xoo}. Fluctuations are assumed to be in the plane perpendicular to $\underline{B}_o$, the applied magnetic field, and fluctuations in $\underline{B}_o$ are ignored (low R_m approximation).

When applied to an MHD Faraday-type generator, E'_{xoo} is the maximum DC Hall field that can be developed in the presence of AC fluctuations (other losses being neglected here), and J_{xoo} is the net Hall current, which must vanish in that case.

Application of these results to special cases is discussed in the recent paper.[8] In particular, it is shown how these effects limit the maximum amplitude of fluctuations in an MHD generator subject to a Velikhov instability, and how these fluctuations can be eliminated or controlled by external circuitry.

<div align="right">J. E. McCune</div>

References

1. J. K. Wright, Proc. Phys. Soc. (London) <u>81</u>, 498 (1963).

2. E. P. Velikhov, Hall Instability of Current-Carrying Slightly Ionized Plasmas, Symposium on Magnetoplasmadynamics, Electrical Power Generation, Newcastle-upon-Tyne, England, September 6-8, 1962.

3. J. E. McCune, Wave Growth and Instability in Partially Ionized Gases, Paper No. 33, International Symposium on MHD Electrical Power Generation, Paris, 1964.

4. J. L. Kerrebrock, AIAA J. <u>2</u>, 1072 (1964).

5. J. E. McCune, Linear Theory of an MHD Oscillator, Research Report RR 198, AVCO-Everett Research Laboratory, 1964.

6. J. F. Louis, J. Lothrop, and T. R. Brogan, Phys. Fluids <u>7</u>, 362 (1964).

7. J. Klepeis and R. J. Rosa, Experimental Studies of Strong Hall Effects and V×B Induced Ionization, Research Report RR-177, AVCO-Everett Research Laboratory, 1964.

8. J. E. McCune, Non-Linear Effects of Fluctuations on MHD Performance, paper to be presented at the Sixth Symposium on Engineering Aspects of Magnetohydrodynamics, Pittsburgh, Pennsylvania, April, 1965.

9. R. J. Rosa, Phys. Fluids <u>5</u>, 1081 (1962).

D. HARTMANN FLOW FRICTION FACTORS — PRESENT AND FUTURE

Considerable time and effort have been expended on the measurement of friction factors in laboratory approximations to Hartmann flow. The purpose of this report is to state what may reasonably be expected from work now under way[1] or planned. There are two main points of interest: the anomalous minimum in f versus M/R curves, and the behavior of f for large M and R. Here, f denotes the friction factor, R the Reynolds number, and M the Hartmann number appropriate to MHD channel flow.

The curves given by Murgatroyd[2] (Figs. XII-6 and XII-7) show that the minimum may disappear at $R \approx 1.2 \times 10^5$. The data of Brouillette and Lykoudis show clearly the
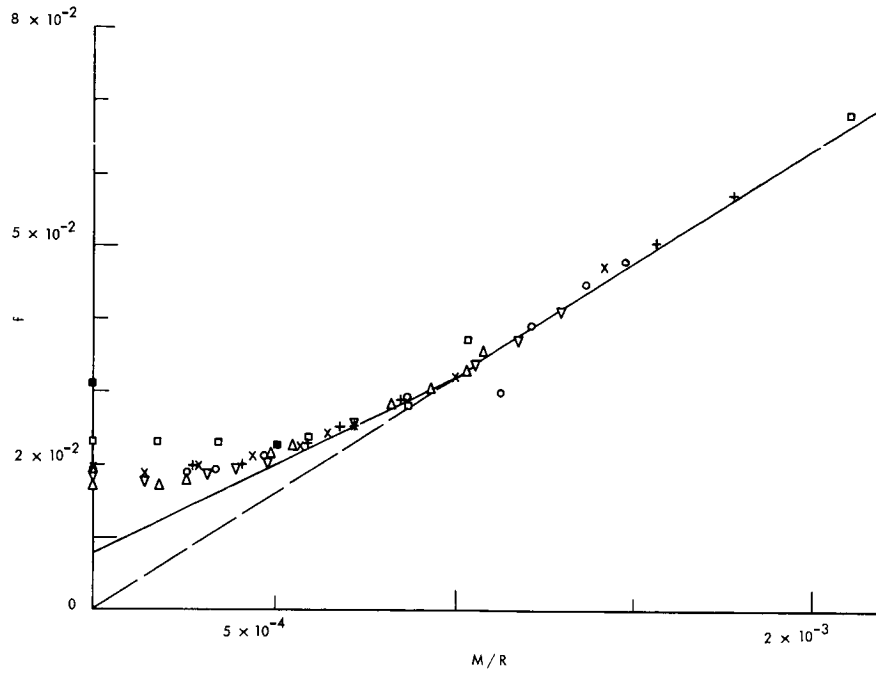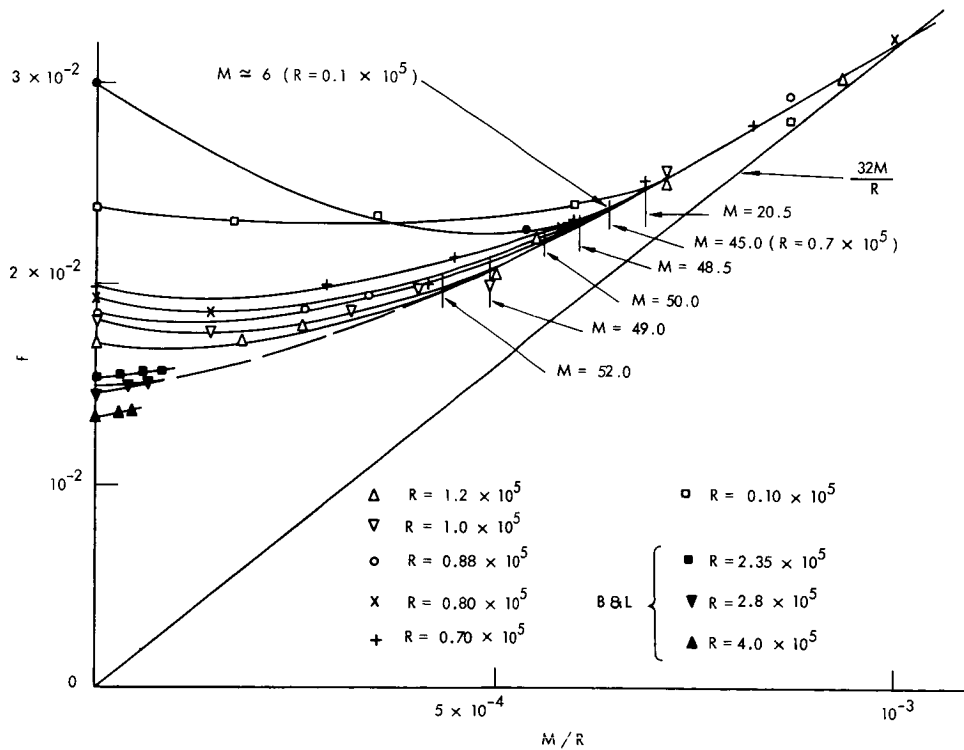
Fig. XII-6. f vs M/R from Murgatroyd's curves.

Fig. XII-7. f vs M/R for data of Murgatroyd and of
Brouillette and Lykoudis.

disappearance of this minimum between $R = 1.1 \times 10^5$ and $R = 2.3 \times 10^5$. The latter set of experiments was not conclusive. Harris[4] stated that the minimum occurred at a constant value of $M^2/R*$, which is approximately the line shown in Fig. XII-8. His analysis was based primarily on the data of Murgatroyd, and his conclusion agrees with all of these data very well, with the exception of the curve for highest R ($R = 1.2 \times 10^5$). This one exception (if valid) may be very important. If the trajectory of the minimum in f in the M-R plane (see Fig. XII-8) does turn down for some R, then the minimum may indeed disappear for some R ($R > 10^5$). The data of Brouillette and Lykoudis, which were not available to Harris, show this turning-down tendency more definitely, although the actual values of M and R are different from Murgatroyd's. Thus, it is possible that an inverted U curve of the type drawn in Fig. XII-8 is the true location of the minimum.
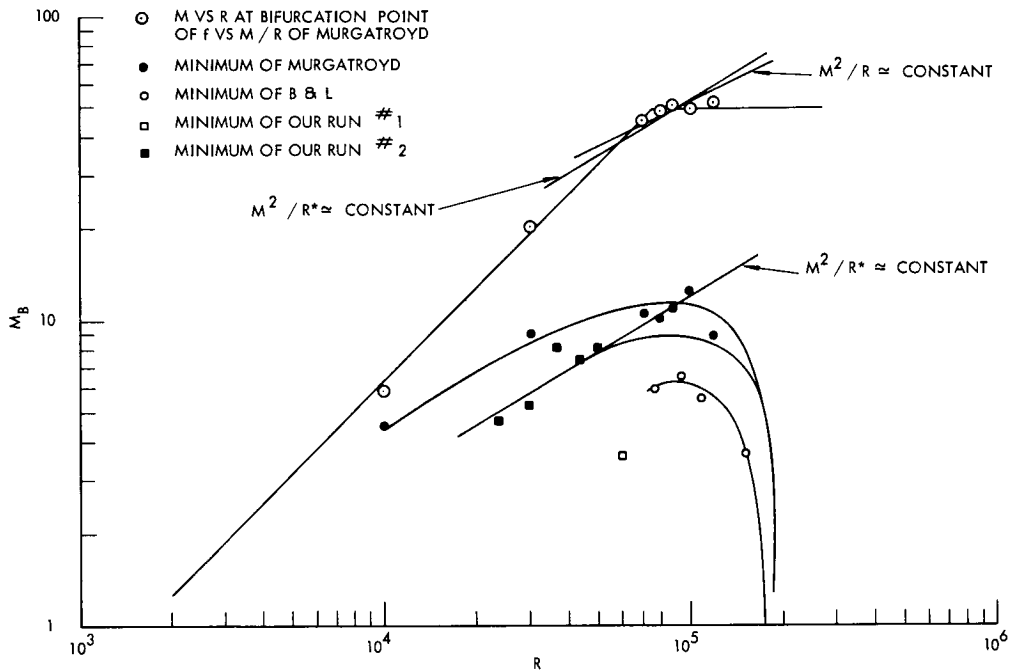

Fig. XII-8. Breakaway value of M vs R.

From the curves of f versus M/R of Murgatroyd (Fig. XII-6) it may be inferred that $f = 32$ M/R for M/R $\geq 10^{-3}$. On a larger scale (Fig. XII-7) the plot shows that an envelope exists for M/R $< 10^{-3}$, from which curves of f break away at different values of M/R, the value depending on the value of R for which the f vs M/R curve is drawn. On the larger scale plot, the values of M (M_B) at which breakaway occurs are given. For $R \geq 7$, 10^4, $M_B \approx 50$. The actual values are plotted in Fig. XII-8, and the dependence is found.

$$M_B/R \approx 6.5 \times 10^{-4}, \qquad R \leqslant 7.5 \times 10^4;$$

$$M_B \approx 50, \qquad\qquad 7.5 \times 10^4 \leqslant R \leqslant 1.2 \times 10^5.$$

For $R > 7.5 \times 10^4$, the point at which f begins to be correlated with M/R, generally known as the parameter characterizing laminar flow, is at constant M. This supports the notion that the transition to laminar flow is governed by the boundary-layer thickness. In truly laminar flow, the velocity profile, and also the boundary layer thickness, are functions only of M.

If we extend the envelope assumption, a curve consistent with those drawn, having

$$\left.\frac{\partial f}{\partial M}\right|_{M=0} = 0$$

might be as drawn in Fig. XII-9. This curve has $f(M=0) = 1.5 \times 10^{-2}$, which corresponds to $R \approx 2 \times 10^5$ on standard ordinary hydrodynamic Moody diagrams (f vs R). Thus we might conclude that the anomalous minimum will not be present for $R > 2.0 \times 10^5$.
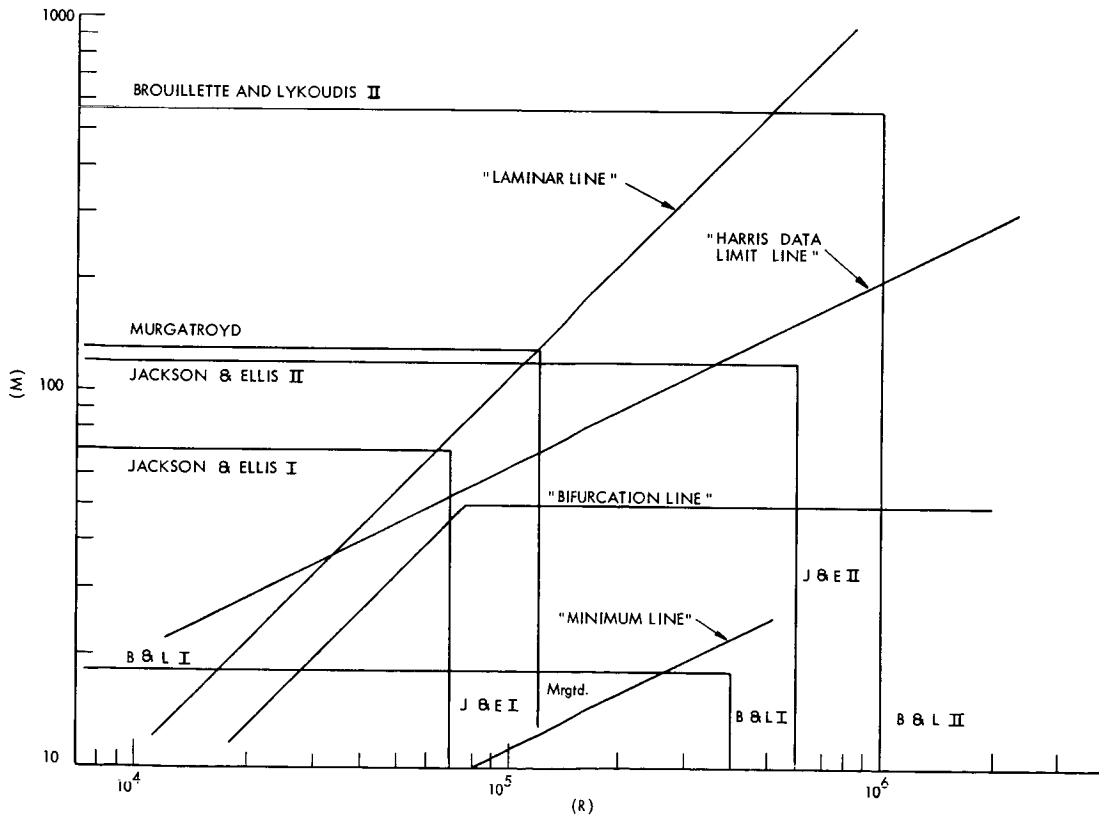

Fig. XII-9. Dependence of f on M^2/R and M/R.

Consider the range of parameters $M^2/R > 4 \times 10^{-2}$, $M/R < 10^{-3}$. Harris predicted that data in this range would be correlated primarily by $M^2/R*$. The breakaway theory, however, predicts that all values of f above and to the left of the "bifurcation line" will be functions of M/R alone. It can be seen that the wedge-shaped region given by the range of parameters considered is completely within the region in which f is a function of M/R alone. Thus data in the range considered may be governed by M/R, rather than by M^2/R or $M^2/R*$.

We propose that the well-known minimum in f versus M/R curves should disappear for $R \gtrsim 2.0 \times 10^5$, provided M^2/R is large. Small values of M/R give little information that cannot be predicted from data now available. If $M > 50$, then any f should be a unique function of M/R in the ranges not yet explored experimentally.

J. R. Ellis, Jr., W. D. Jackson

References

1. W. D. Jackson and J. R. Ellis, Jr., Friction-factor measurements in liquid-metal magnetohydrodynamic channel flows, Quarterly Progress Report No. 73, Research Laboratory of Electronics, M.I.T., April 15, 1964, pp. 122-125.

2. W. Murgatroyd, Experiments on magneto-hydrodynamic channel flow, Phil. Mag. 44, 1348 (1953).

3. E. G. Brouillette and P. S. Lykoudis, Measurements of skin friction for turbulent magnetofluidmechanic channel flow, Purdue Research Foundation Project No. 3093, August 1962.

4. L. P. Harris, Hydromagnetic Channel Flows (Technology Press of M.I.T., Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1960).

E. MAGNETOHYDRODYNAMIC INDUCTION MACHINE WITH LAMINAR FLUID FLOW

In our previous analysis of the MHD induction machine only constant fluid velocity has been considered in illustrating the principles of operation.[1,2] The fluid velocity profile and viscosity will have important effects for two reasons: (i) the velocity profile may drastically alter the power flow; and (ii) the viscous losses may be large. An exact consideration of profile effects is impossible, but a solution can be obtained for laminar flow with reasonable approximations.

The laminar-flow solution gives insight into the mutual interaction. The flow in a practical machine, however, will probably be highly turbulent, because of the high velocities or (hydraulic) Reynolds numbers R_e required to achieve reasonable power densities. Turbulence in induction-driven flows is not fully understood, so that the limits on R_e for laminar flow as a function of the applied magnetic field and flow parameters are unknown. A qualitative idea of turbulence, the turbulent velocity profile, and the effect of

turbulence on machine performance has been obtained,[3] and will be presented in a latter report.

1. Model and Equations

The model to be analyzed is shown in Fig. XII-10. The fluid flows in the x direction between two parallel exciting plates of infinite extent in the x and z directions, spaced 2a apart. The region outside the plates is filled with a core of permeability μ_c and conductivity σ_c. The exciting plates, separated from the fluid and core by insulators of infinitesimal thickness to prevent current flow in the y direction, are assumed to be thin so that they can be replaced by current sheets. The plates are driven by a current source that gives an even or symmetric surface current density

$$\vec{K} = \vec{i}_z NI \cos (\omega t - kx); \tag{1}$$

which represents a traveling current wave of amplitude NI, frequency ω, wavelength $\lambda = \frac{2\pi}{k}$, and velocity $v_s = \frac{\omega}{k}$. The machine of finite length has been treated previously for a constant fluid velocity.[4,5] The extension to include both velocity profiles and finite length is possible for a slit-channel machine because the field solution is independent of the velocity profile for this case only.

The electromagnetic fields are determined from Maxwell's equations with the usual MHD approximation of neglecting displacement currents. The analysis is simplified by the use of a vector potential $\vec{A}$ and scalar potential ϕ defined as

$$\vec{B} = \nabla \times \vec{A} \tag{2}$$

and

$$\vec{E} = -\nabla \phi - \frac{\partial \vec{A}}{\partial t}. \tag{3}$$

Noting that Ohm's law in a moving fluids is $\vec{J} = \sigma(\vec{E} + \vec{v} \times \vec{B})$ and substituting Eqs. 2 and 3 in Maxwell's equations give

$$\nabla^2 \vec{A} - \mu\sigma \frac{\partial \vec{A}}{\partial t} + \mu\sigma(\vec{v} \times \nabla \times \vec{A}) = 0, \tag{4}$$

and

$$\nabla^2 \phi - \mu\sigma \frac{\partial \phi}{\partial t} = 0. \tag{5}$$

Here,

$$\nabla \cdot \vec{A} + \mu\sigma\phi = 0 \tag{6}$$

has been chosen to uncouple Eqs. 4 and 5. The vector potential is due solely to currents, so that it, as well as $\vec{J}$, is in the z direction and independent of z, and $\phi = 0$.
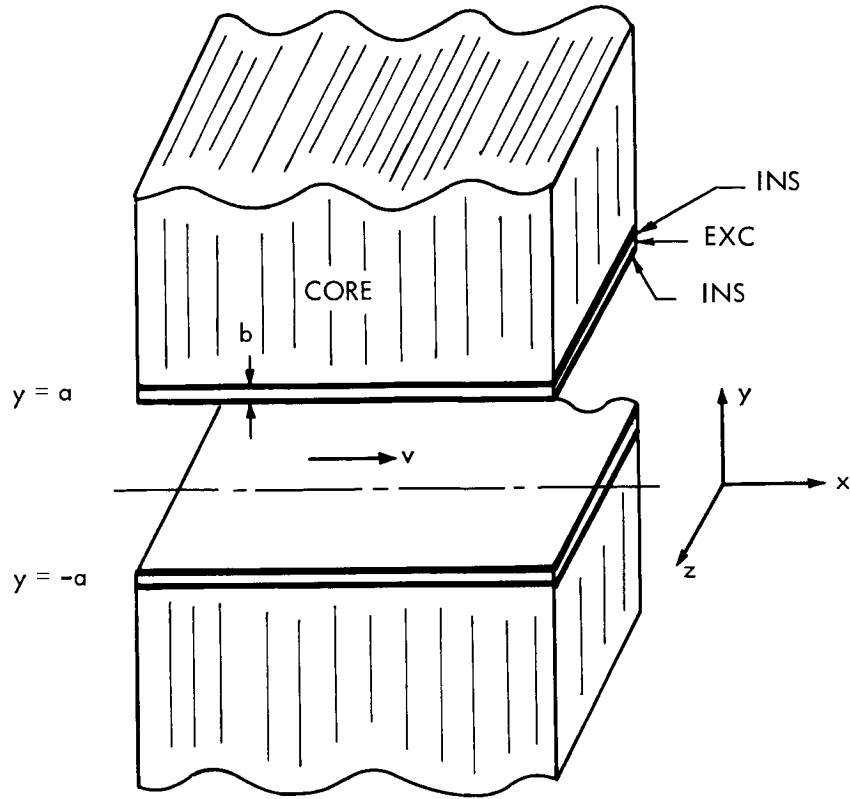
Fig. XII-10. The model.

The fluid behavior is determined from the Navier-Stokes or conservation of momentum equation and the conservation of mass or continuity equation. For an incompressible fluid of mass density ρ and absolute viscosity η, these equations are

$$\rho\left(\frac{\partial}{\partial t}+\vec{v}\cdot\nabla\right)\vec{v} = -\nabla p + \eta\nabla^2\vec{v} + \vec{J}_f \times \vec{B}_f, \tag{7}$$

and

$$\nabla \cdot \vec{v} = 0, \tag{8}$$

where p is the pressure, subscripts f and c denote fluid and core quantities, and $\vec{J}_f$ and $\vec{B}_f$ can be expressed in terms of $\vec{A}_f$.

The equations can be solved only for laminar flow, as little is known about solutions for turbulent flows either with or without a magnetic field. Even then, an analytical solution is clearly impossible, because of the nonlinear terms and the two-directional coupling, so that a reasonable approximation is made — that the fluid velocity is laminar, in the x direction, and independent of x and t. This approximation and the resulting equations are discussed below. An analytical solution of the simplified equations is

obtained for a slit channel. In general, even the simplified equations can only be solved numerically because the nonlinear terms and the two-directional coupling remain. The numerical techniques and results are also discussed.

2. Approximate Equations for Laminar Flow

To solve Eqs. 4 and 7, they must be simplified, because of the product type of non-linearities. Linearization with the constant-velocity solution used as the starting point is not valid for two reasons: the constant velocity does not match the boundary conditions, and the perturbations are not small. A power-series solution in the magnetic Reynolds number $R_M = \mu_f \sigma_f v_s / k$ is not desired because it would be limited to small R_M, which does not include the range of interest for power generation. Writing $\vec{v}$ and $\vec{A}_f$ as Fourier series in $e^{jn(\omega t - kx)}$, because of the simple $e^{j(\omega t - kx)}$ excitation of the current sheets, gives an infinite number of completely coupled equations that still involve products and derivatives with respect to y. These equations are not amenable to solution.

The equations are simplified by assuming the velocity to be completely in the x direction. The continuity equation then requires $\vec{v}$ to be independent of x. The electromagnetic pressure gradient on each strip, $(\vec{J}_f \times \vec{B}_f)_x$, consists of a constant term plus a second harmonic in $(\omega t - kx)$, since $\vec{J}_f$ and $\vec{B}_f$ vary as $e^{j(\omega t - kx)}$ for a velocity independent of x and t. The total force on a strip per wavelength in the x direction is a constant, so that v should not have any time dependence for an incompressible fluid. Thus, $\vec{v} = \vec{i}_z v(y)$ is a consistent and reasonable approximation to the actual flow, provided the transverse fluid velocity is small compared with the velocity along the machine. This can be justified for a slit-channel machine by examining the driving forces.[6]

The equations are put in a more tractable form by using complex notation, $\vec{A}_f(x,y,t) = R_e \left\{ \vec{i}_z \underline{A}_f(y) \, e^{j(\omega t - kx)} \right\}$, and by defining the normalized variables $\tilde{y} = \dfrac{y}{a}$, $u(\tilde{y}) = \dfrac{v(\tilde{y})}{\bar{v}}$, and $\tilde{F}(\tilde{y}) = \dfrac{\underline{A}_f(\tilde{y})}{\underline{A}_{fo}}$, where a is the channel half-height, $\bar{v}$ is the average velocity, and $\underline{A}_{fo}$, the vector potential at the center of the channel, is determined by the boundary conditions. In terms of v_s and the average slip $\bar{s}$, $\bar{v} = (1-\bar{s})v_s$, which is considered to be a constant of the solution.

The Hartmann number for DC flows is

$$M_o^2 = \frac{\sigma_f a^2}{\eta} B_o^2,$$ (9)

where B_o is the transverse magnetic field. If we define an effective Hartmann number $M(\tilde{y})$ for the induction machine,

$$M^2 = \frac{\sigma_f a^2}{\eta} \frac{A_f A_f^*}{2},$$ (10)

in terms of the rms transverse magnetic field, where $a = ak$, the normalized equations become

$$\frac{d^2 F}{d\tilde{y}^2} - a^2[1 + jR_M - jR_M(1-\bar{s})u]\, \underline{F} = 0,$$ (11)

and

$$\frac{d^2 u}{d\tilde{y}^2} - M^2 u = \frac{a^2 p_o}{\eta(1-\bar{s})v_s} - \frac{M^2}{(1-\bar{s})}$$ (12)

from Eq. 4 and the space or time average of Eq. 7. Equation 12 is identical to the Hartmann profile equation for a DC generator in terms of a proper loading factor.

The original set of equations has been simplified to two coupled nonlinear ordinary differential equations. An analytical solution, possible only for a slit channel, is given below. The general case is also treated numerically. The solution could be obtained by series techniques, but these are of limited validity.[7] Also, the present approach is easier to use and more flexible.

The electromagnetic powers can be written in terms of the normalized variables, but the simple power relations[1,8] no longer exist. In general, the magnitudes of P_m and P_r, the mechanical power output and the power dissipated in the fluid because of its finite conductivity, are increased over their constant-velocity values for a generator, because of the circulating currents that are set up, since the velocity drops below synchronous speed near the walls. This region acts as a pump, and the net behavior of the machine may be changed from a generator to a pump or damper.

3. Slit-Channel Solution

For a slit channel, $\gamma a \ll 1$, the equations can be simplified to yield analytical solutions. The electromagnetic field and the vector potential are independent of y, and the total current in the field at a given value of x is therefore independent of the velocity profile. This means that the vector potential can be determined first, and then the profile and powers. The vector potential, depending only on the total current, is the same as the slit-channel field that is found for a constant velocity,

$$A_f = \frac{\mu_f NI}{k(\gamma^2 a + \kappa \delta)},$$ (13)

where $\gamma^2 = 1 + jsR_M$, $\kappa = \frac{\mu_f}{\mu_c}$, and $\delta^2 = 1 + j\frac{\mu_c \sigma_c v_s}{k}$.

The velocity, obtained from Eq. 12 for M constant, is the Hartmann profile,

$$u(\tilde{y}) = \frac{\left(1 - \frac{\cosh M\tilde{y}}{\cosh M}\right)}{\left(1 - \frac{\tanh M}{M}\right)};$$ (14)

so that, for a slit channel and laminar flow, the profile for the induction machine is identical to that for a DC machine with a Hartmann number based on the rms transverse magnetic field. The Hartmann profile is plotted in Fig. XII-11 for several values of M. The curve for M = 0 is the parabolic profile of laminar hydrodynamic flow.

The powers for the slit-channel machine[1,8] are

$$P_s = P_0 \frac{a\bar{s}R_M}{(\gamma^2 a + \kappa\delta)(\gamma^{*2}a + \kappa\delta^*)},$$ (15)

$$P_m = (1-\bar{s})P_s F_m(u),$$ (16)

and

$$P_r = \bar{s}P_s F_r(u),$$ (17)

where $P_0 = \mu_f v_s N^2 I^2 c\ell$, and c and ℓ are the machine length and width. Here P_s, the time-average real power supplied to the fluid, is independent of the profile, since it depends only on the fields. The equations are identical with the results for constant fluid velocity except for the profile factors

$$F_m(u) = \frac{1}{\bar{s}}[1-(1-\bar{s})\overline{u^2}],$$ (18)

and

$$F_r(u) = \frac{1}{\bar{s}^2}[2\bar{s}-1+(1-\bar{s})^2\overline{u^2}],$$ (19)

where

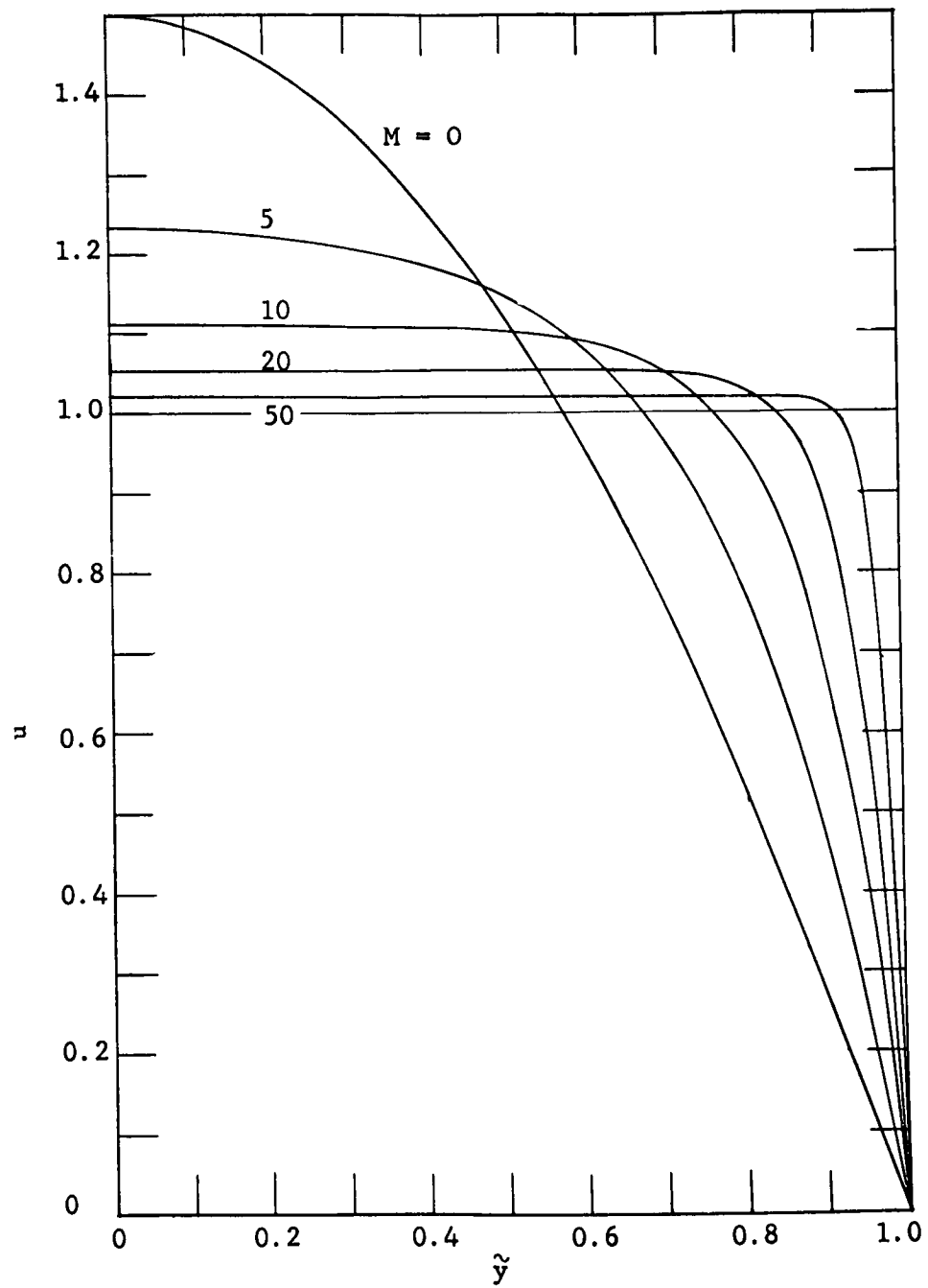$$\overline{u^2} = \int_0^1 (u(\tilde{y}))^2 \, d\tilde{y},$$ (20)

Fig. XII-11. Normalized Hartmann velocity profiles.

the average of the velocity squared, is always ≥ 1. For a generator, F_m and F_r are ≥ 1, and the $I^2 R$ losses are increased for the same power output. For a pump, $F_m \leq 1$.
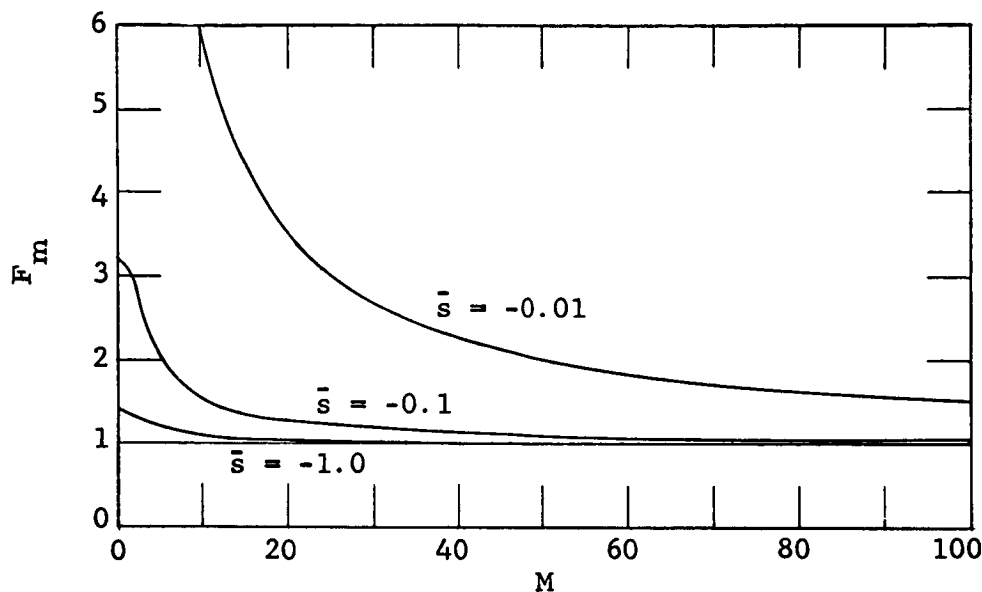

Fig. XII-12. F_m for Hartmann profile.

The profile factors for a Hartmann profile, plotted in Figs. XII-12 and XII-13 for several negative values for $\bar{s}$, show that the factors increase as $|\bar{s}|$ decreases and as the profile deviates from a constant. This is to be expected for a generator because, as $\bar{s}$ approaches zero or as the profile becomes more rounded, the size and relative importance of the positive-slip region near the wall increases. In this region the machine is acting as a pump, so that the losses resulting from circulating currents are increased.

The generator efficiency, P_s/P_m, is

$$e_g = \frac{1}{(1-\bar{s})F_m},$$ (21)

since P_s is unchanged but the input power is increased by F_m. At $\bar{s} = 0$, there is no power output, but the circulating currents still exist, there is power input, and the efficiency is zero. There is a peak in the efficiency, so that decreasing $|\bar{s}|$ further results in a poorer efficiency because more of the fluid is pumped, in contrast to the constant-velocity case for which e_g approached one as $\bar{s}$ approaches zero. Curves of e_g versus $\bar{s}$ for several values of M are shown in Fig. XII-14.

These results, derived by assuming $\gamma a \ll 1$, are valid for a much larger range than
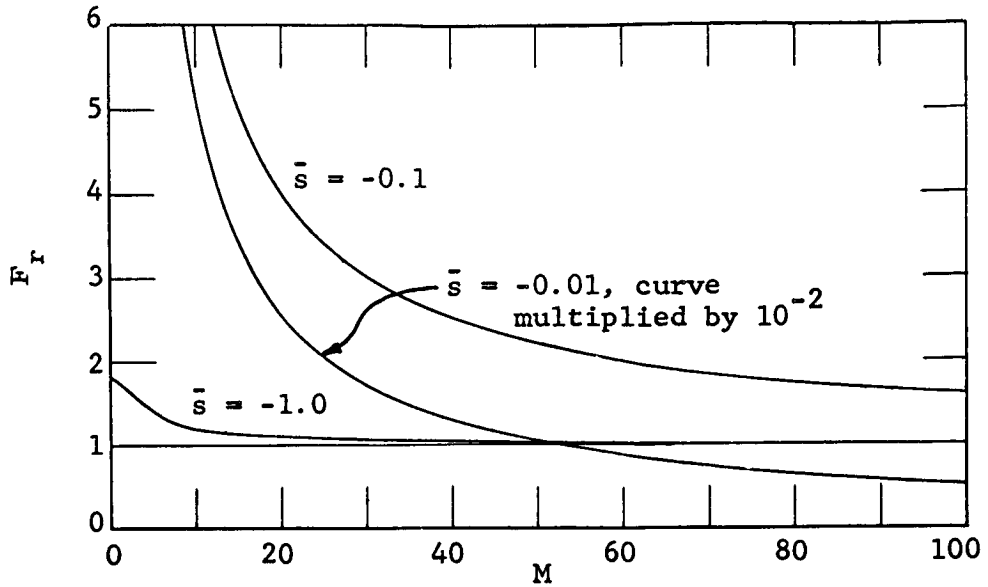
Fig. XII-13. F_r for Hartmann profile.

expected. Comparison with the numerical results shows good agreement for a as large as 0.1; that is, errors of a few per cent for $a = 0.1$ and $|\bar{s}R_M|$ up to 100.
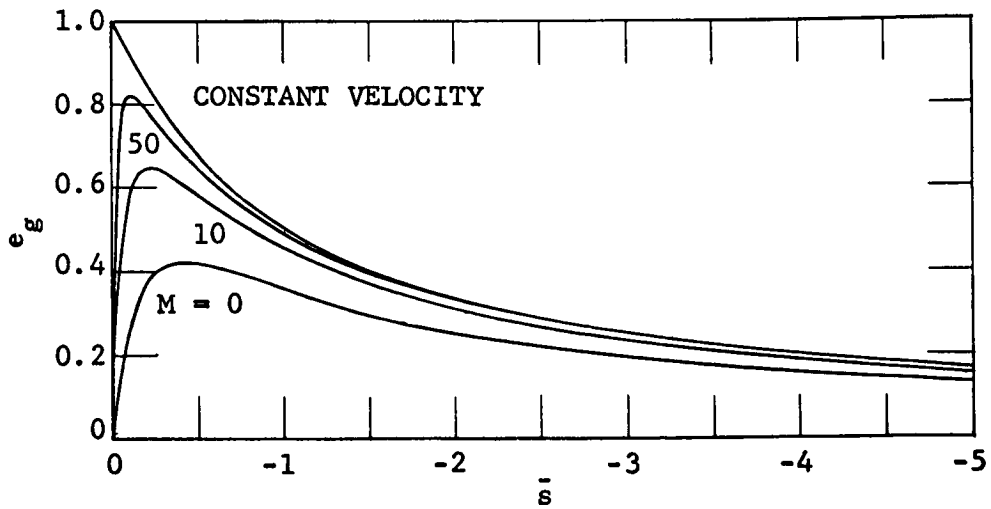

Fig. XII-14. Generator efficiency for slit-channel machine with Hartmann profile.

4. An Iterative Solution

The vector potential in the fluid can be determined from Eq. 11 when the velocity is specified. The result when the profile for ordinary hydrodynamic flow, either laminar

or turbulent, is used is the approximate solution when the electromagnetic force is small compared with the viscous force. Otherwise, the result may not correspond to an actual flow, but still gives information on the dependence of the fields and powers on the velocity profile. Equation 11 is a linear homogeneous second-order differential equation for $\tilde{F}(\bar{y})$ in terms of $u(\bar{y})$ and the machine parameters. Several numerical methods were used to solve Eq. 11 when the velocity was specified, and the preferred method was chosen by testing on the case $u(\bar{y}) = 1$, for which the exact solution is known. The numerical integration procedure for Eq. 11 worked quite well.[9]

Equation 12 can be solved for the fluid velocity profile when the vector potential is specified. The profile obtained is the small sR_M solution if the field with no fluid present is used, corresponding to the case for which the field is not appreciably affected by the fluid. The solution of the nonhomogeneous Eq. 12 is more complicated than the solution of the homogeneous Eq. 11. Before, there was only a homogeneous solution for specified initial conditions, which was scaled to match the boundary conditions. Now, there is a homogeneous solution and one or two particular solutions, which depend on the approach used. Initial conditions are specified, and then a linear combination of the solutions is used to match the boundary conditions. Several approaches to solving Eq. 11 were tested. Difficulty occurred, because of the exponential-like behavior of the solutions, which could not be avoided. This puts a limit on the parameters for which the solution can be obtained.[10]

The techniques developed for solving Eqs. 11 and 12 can be combined to obtain an exact solution for laminar flow by iterating. The electromagnetic field for a constant fluid velocity is used as the starting point, and Eqs. 12 and 11 are solved repetitively for the new velocity and field in that order, until the solution converges to the desired accuracy. The convergence is good for R_M small, but becomes worse as R_M increases because the fluid profile has more effect on the field.

All the $a = 0.1$ cases tested with the iterative procedure checked with the slit-channel results, except for $sR_M = -250$, where M^2 varied by a factor of three across the channel, and the profile was no longer Hartmann. The small γa case is excluded from further consideration here as it is better treated by the methods discussed in section 3.

Some of the results for large γa are given in Figs. XII-15, XII-16, XII-17, and in Table XII-1. The number of iterations required for convergence to five figures is given in Table XII-1, where > means that the result is close, but more than the number tried were required. The number increases with increasing $\bar{s}$ and R_M. The excitation magnitude is contained in the normalization constant $FOR = \mu_f(NI)^2 a/2\eta v_s$.

The velocity profiles for $a = 1$, $R_M = 1$ and 10, are plotted in Figs. XII-15 and XII-16 for $\bar{s} = -0.1$, -1, and -10. The two $\bar{s} = -0.1$ curves look similar, but are quite different near the walls. The power density and efficiency for $\bar{s} = -10$ are greater than for a
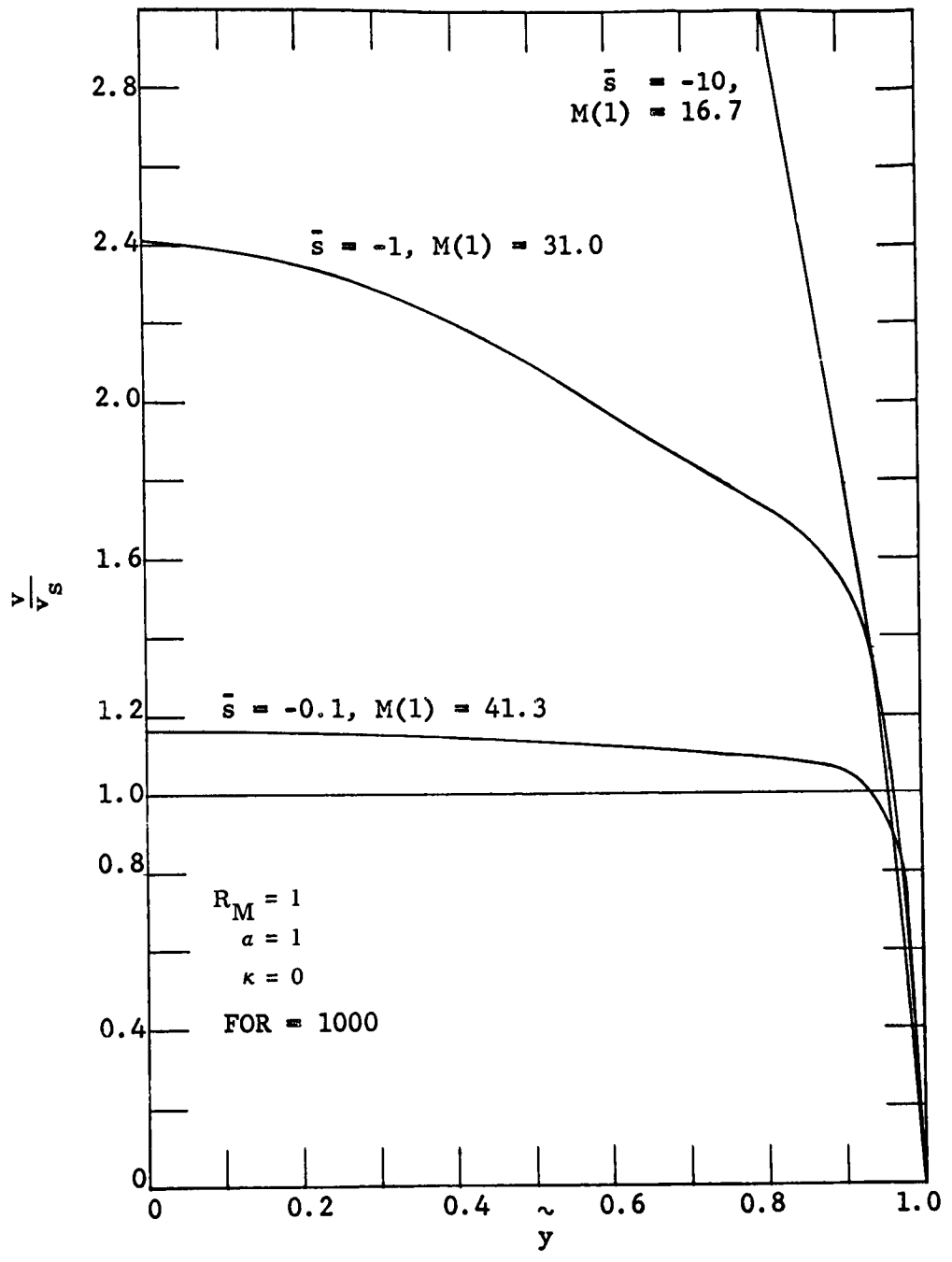
Fig. XII-15. Iterative solution for velocity, $a = 1$,
$R_M = 1$, FOR = 1000.

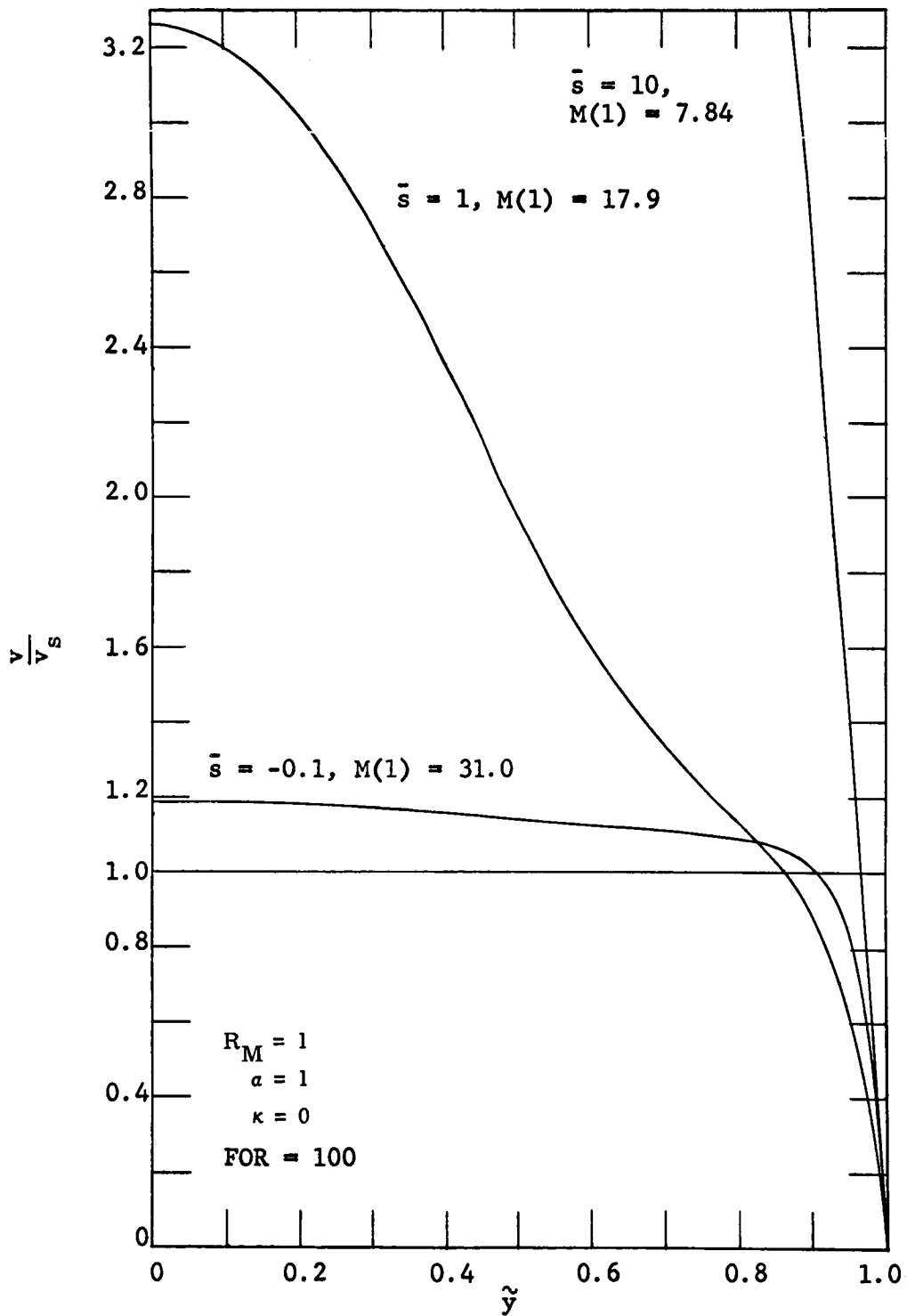Fig. XII-16. Iterative solution for velocity, $a = 1$, $R_M = 10$, FOR = 100.

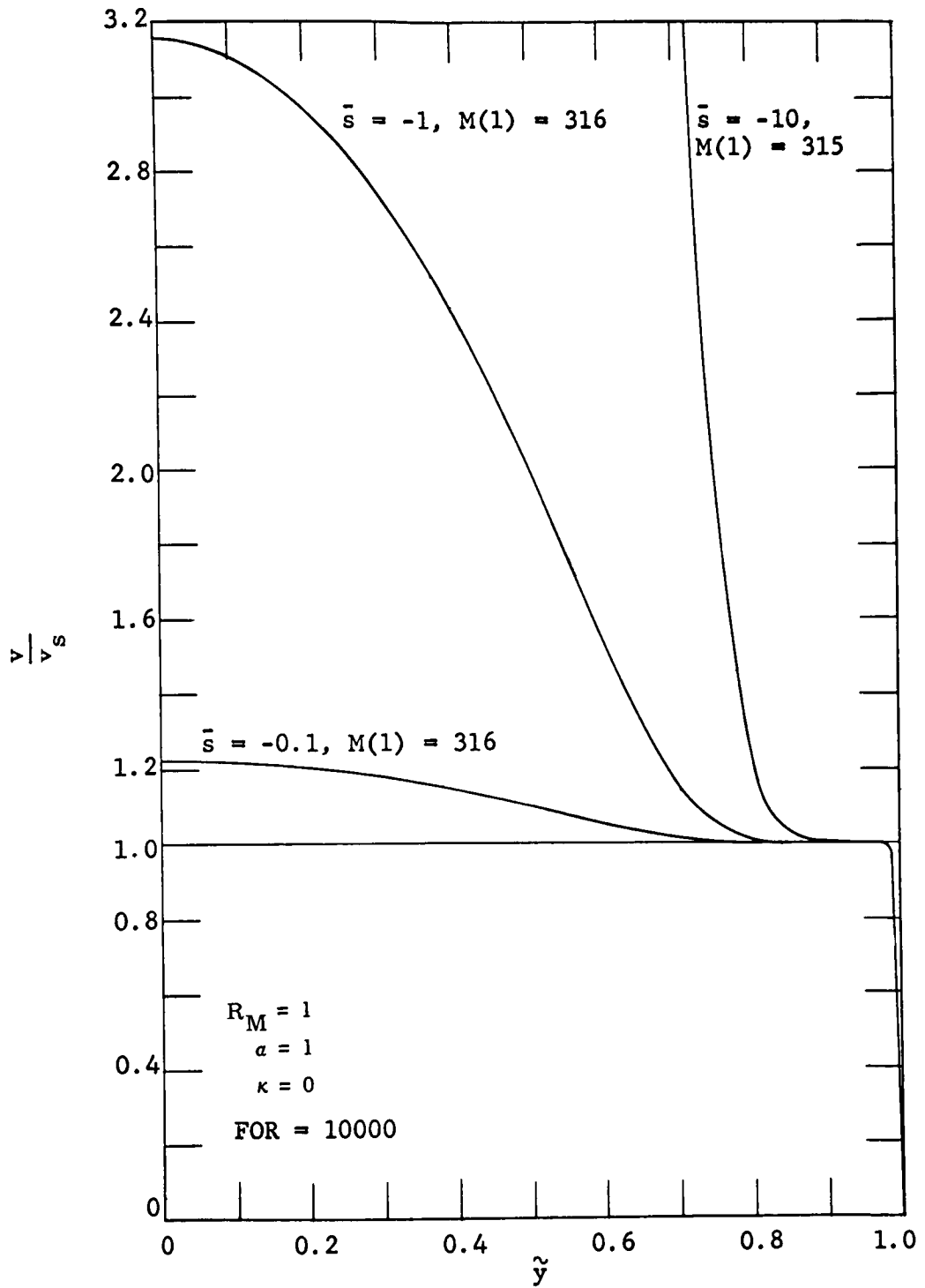Fig. XII-17. Iterative solution for velocity, $a = 10$, $R_M = 1$, FOR = 10,000.

constant fluid velocity because the average slip seen by the field is less than the average for the fluid. This is not of practical significance because the power density and efficiency are both low.

The profiles for $a = 10$ and $R_M = 1$, show the strong field influence at the wall. The electromagnetic dominance is more pronounced for larger excitation and extends farther

Table XII-1.   Iterative solution results, $\kappa = 0$.

$\overline{s}$	R_M	a	FOR	$M(1)$	$\dfrac{P_s}{P_o}$	e_g, %	Required number of iterations
-0.1	1	1	1000	41.3	-0.0753	69.9	2
-1				31.0	-0.456	47.1	4
-10				16.7	-0.168	10.3	>5
-0.1	10	1	100	31.0	-0.287	57.3	>3
-1				17.9	-0.0388	13.8	>5
-10				7.84	-0.0957	17.8	>5
-0.1	1	10	1000	99.5	0.0937	Pump	2
-1				99.5	0.0885	Pump	5
-10				97.2	0.0443	Damper	>6
-0.1	1	10	10,000	316	0.0310	Pump	3
-1				316	0.0302	Pump	5
-10				315	0.0238	Damper	>5

into the fluid. None of the tested $a = 10$ cases will operate as a generator, since the pumping of the boundary layer dominates, as shown in Table XII-1.

The exact solution has been obtained for several sets of parameters, and can be extended to others if desired. One important result is to eliminate the large γa machine from further consideration. Not only is the power density low, but it will not operate as a generator. The slit-channel results have been used to obtain predicted performance characteristics.[11]

E. S. Pierson, W. D. Jackson

References

1. E. S. Pierson, Power flow in the magnetohydrodynamic induction machine, Quarterly Progress Report No. 68, Research Laboratory of Electronics, M.I.T., January 15, 1963, pp. 113-119.

2. W. D. Jackson, E. S. Pierson, and R. P. Porter, Design Considerations for MHD Induction Generators, International Symposium on Magnetohydrodynamic Electrical Power Generation, Paris, July 6-11, 1964.

3. E. S. Pierson, The MHD Induction Machine, Sc. D. Thesis, Department of Electrical Engineering, M. I. T., Cambridge, Massachusetts, 1964; see Chap. 5.

4. Ibid., Chap. 6.

5. E. S. Pierson and W. D. Jackson, Magnetohydrodynamic induction machine of finite length, Quarterly Progress Report No. 75, Research Laboratory of Electronics, M. I. T., October 15, 1964, pp. 92-103.

6. E. S. Pierson, Sc. D. Thesis, op. cit., pp. 76-78.

7. J. P. Penhune, Energy Conversion in Laminar Magnetohydrodynamic Channel Flow, ASD Technical Report 61-294, Research Laboratory of Electronics, M. I. T., August 1961, pp. 75-78.

8. E. S. Pierson, The MHD Induction Machine, Sc. D. Thesis, op. cit., Sec. 3. 3.

9. Ibid., Sec. 4. 4.

10. Ibid., Sec. 4. 5.

11. E. S. Pierson and W. D. Jackson, Prediction of magnetohydrodynamic induction generator performance, Quarterly Progress Report No. 76, Research Laboratory of Electronics, M. I. T., January 15, 1965, pp. 156-164.

F. BEHAVIOR OF DRY POTASSIUM VAPOR IN ELECTRIC AND MAGNETIC FIELDS

A description of preliminary measurements of the electrical properties of wet and dry potassium vapor was given in Quarterly Progress Report No. 74 (pages 155-166). These results have also been reported in detail.[1]

Although the results for nonequilibrium conduction in wet vapor appeared to be consistent with theoretical predictions, those for the dry vapor were confusing. In particular, the measured electrical conductivities in the nonequilibrium regime were as much as a factor of ten larger than those expected from theory.

More refined measurements of the conductivity of the dry vapor have now been carried out, both with and without a magnetic field. The present discussion is a preliminary report of these results.

1. Apparatus

The most important modifications to the experimental facility are indicated schematically in Fig.XII-18. Whereas in the earlier experiments[1] the best section was located in the condenser and electrical connections were taken out through the condenser, the test section is now located in an argon-purged vacuum enclosure below the condenser. This new arrangement has the advantage of making the test leads more accessible and perhaps less susceptible to shorting. It has the disadvantage that many joints between the electrical insulator and the metal parts must be made gas tight.

In the present design the boron-nitride test-section tube is joined to the metal parts by tapered joints as shown in Fig. XII-18. The probes are held in stainless-steel taper pins. This design was not as leak-tight as we desired, but it was adequate to produce some data.

An electric field was imposed on the plasma by means of the electrodes at the top and bottom of the test section.   A magnetic field was imposed in a direction perpendicular to the tube axis (and to the electric field).   This combination leads to a Hall field
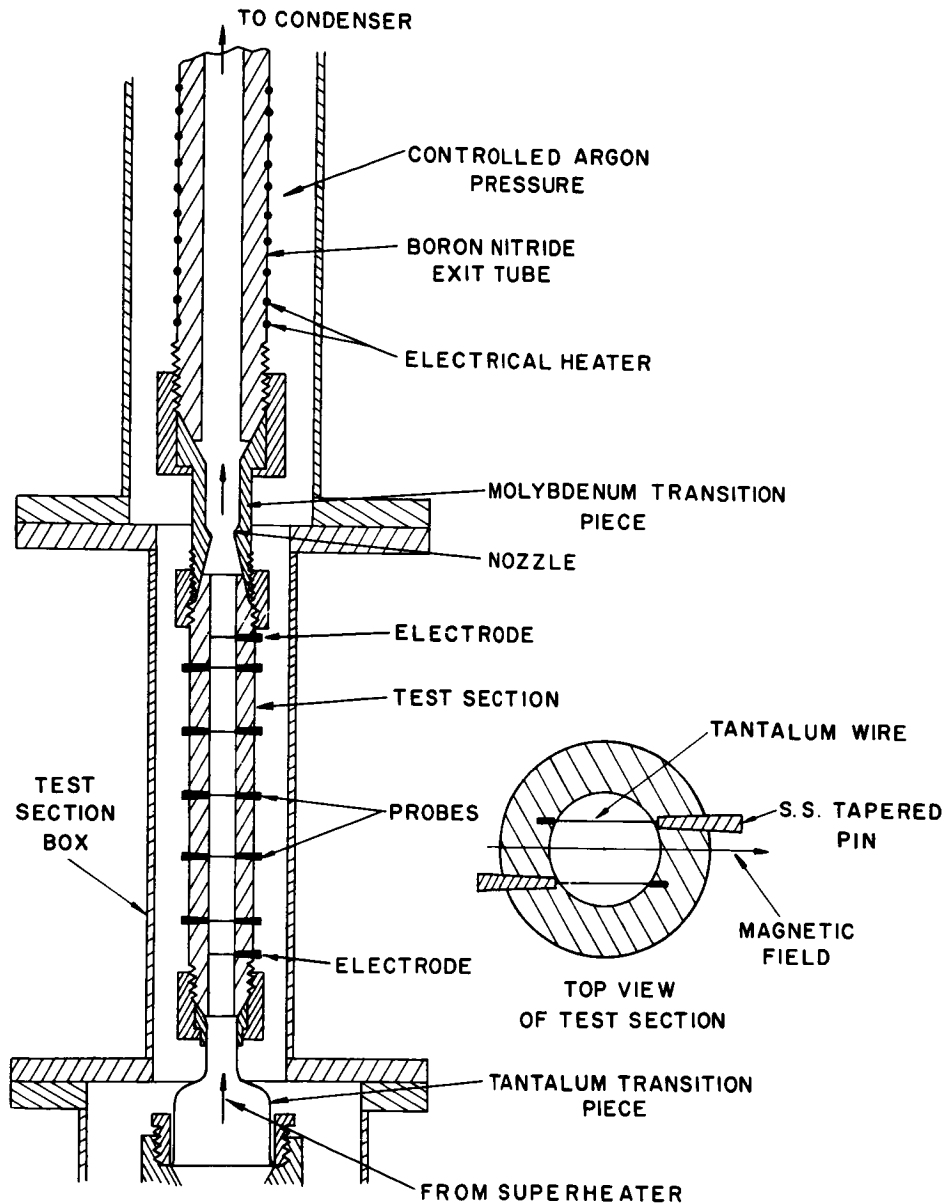

Fig. XII-18. Test section assembly.

normal to both, which was measured by means of the pairs of probes shown schematically at the right in Fig. XII-18. In order to prevent shorting of the test section through ground at its two ends, a long (approximately 1 ft) insulating exit tube was placed between

the test section and condenser. Both the test section and the exit tube were heated to prevent condensation.

A throat, located downstream of the test section, was sized to give a Mach number of 0. 6 in the test section.

2. Experimental Results

The experimental results consist of axial and transverse voltages, as functions of the axial current and the magnetic field.

In Fig. XII-19 the variation of voltage along the test section is given as a function of the current density, for zero magnetic field. The curves are linear; this indicates that
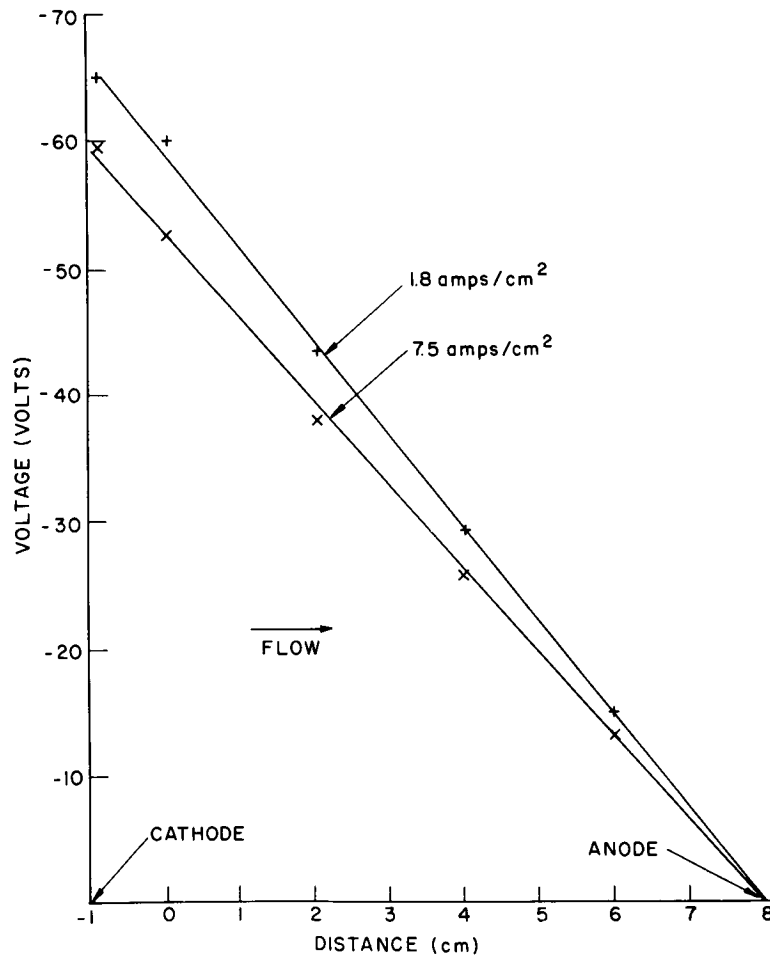

Fig. XII-19. Voltage distribution along channel for various current densities, and zero magnetic field.

a nearly constant conductivity was attained in the plasma.

From these and other similar results, the plasma conductivity was determined as a

function of the current density. The results are compared with the two-temperature theory[2] in Fig. XII-20. Two theoretical curves are given to indicate the uncertainty in plasma pressure and temperature. The spread of the data is indicated by the bars, and the mean of several point by the circles. We conclude that, within the uncertainty in the plasma conditions, the measured conductivity agrees with that predicted theoretically. In the light of current understanding of two-temperature plasmas, this may be interpreted to mean that the electron collision cross section of potassium is 250×10^{-16} cm^2 within approximately 15 per cent.
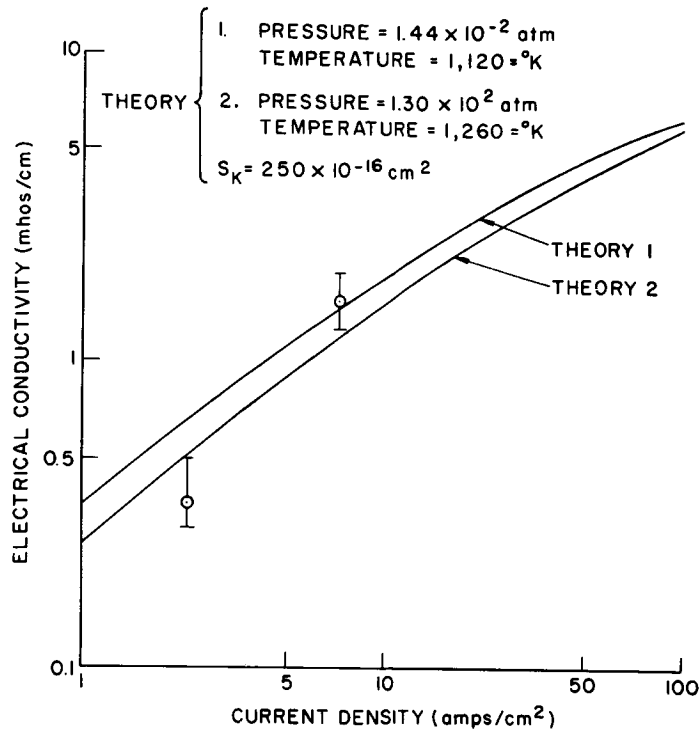

Fig. XII-20. Nonequilibrium electrical conductivity of pure superheated potassium vapor.

Further confirmation of this value is offered by the measurements of Hall voltage. We note first of all that if the axial and transverse directions are denoted z and y, respectively, while the magnetic field is in the x direction, the current densities are given by

$$j_z = \sigma E_z + \beta j_y \tag{1}$$

$$j_y = \sigma E_y + \beta(j_0 - j_z), \tag{2}$$

where σ is the local conductivity, β is the local Hall parameter, and $j_0 = e n_e V$, with n_e being the local electron density and V the flow velocity.

Now, if $j_y = 0$, then $j_z = \sigma E_z$, and

$$\frac{E_y}{E_z} = \beta \left(1 - \frac{j_0}{j_z} \right). \tag{3}$$

Having j_0 and j_z, we can therefore deduce a value of β from the measured E_y and E_z. Such results are shown as the triangular points in Fig. XII-20. There is a great deal of scatter, and considerable systematic deviation from the theoretical result based on the cross section of 250×10^{-16} cm^2.

In fact, j_y is not exactly zero, because of local shorting of the Hall voltage by non-uniformities and by the probes. But if we define an effective conductivity,

$$\sigma_{eff} = \frac{j_z}{E_z}, \tag{4}$$

we have from Eq. 1,

$$1 = \frac{\sigma}{\sigma_{eff}} + \beta \frac{j_y}{j_z} = \frac{\sigma}{\sigma_{eff}} + \beta \left[\frac{\sigma}{\sigma_{eff}} \frac{E_y}{E_z} - \beta \left(1 - \frac{j_0}{j_z} \right) \right].$$

Let $\beta_{eff} = E_y/E_z (1 - j_0/j_z)$, and set $s = \sigma_{eff}/\sigma$ and $b = \beta/\beta_{eff}$. Then

$$1 - \frac{1}{s} = \beta \left(1 - \frac{j_0}{j_z} \right) \left[\frac{\beta_{eff}}{s} - \beta \right] = \beta_{eff}^2 \left(1 - \frac{j_0}{j_z} \right) \left[\frac{b}{s} - b^2 \right].$$

Solving for b, we find

$$b = \left\{ \frac{1}{2} + \sqrt{ \frac{1}{4} - \frac{(s-1)s}{\left(1 - \frac{j_0}{j_z} \right) \beta_{eff}^2} } \right\} \frac{1}{s}.$$

Now taking $s = \sigma_{eff}/\sigma$ as the ratio of σ_{eff} to the σ determined experimentally for B = 0, and evaluating $1 - j_0/j_z$ similarly from the experimental results for B = 0, we can compute b, and hence the actual value of β. The data corrected in this way are shown as the circles in Fig. XII-21. They agree quite well with the theoretical results. We conclude
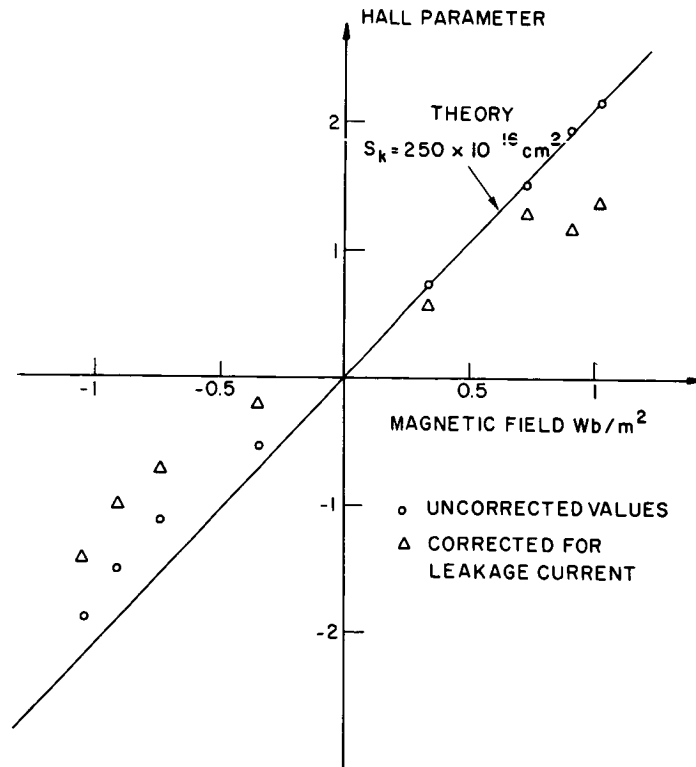
Fig. XII-21. Hall parameter versus magnetic field. (Triangles represent the measured Hall field; circles are corrected for Hall field shorting.)

that β and σ are essentially constant, and that the deviation of the measured Hall fields from the expected value is due to shorting of the Hall fields by the probes, by the inlet and outlet nozzles or, perhaps, by the walls.

3. Conclusions

To summarize, we conclude that, within the accuracy of this experiment, the electric conductivity of dry potassium vapor is as predicted by the two-temperature theory.[2]

The conductivity measurements for zero magnetic field and Hall field measurements at Hall parameters up to 2 indicate independently an electron collision cross section of 250×10^{-16} within 15 per cent.

J. L. Kerrebrock, M. A. Hoffman, A. Solbes

References

1. A. W. Rowe and J. L. Kerrebrock, Nonequilibrium Electric Conductivity of Wet and Dry Potassium Vapor, Technical Documentary Report No. APL-TDR-64-106, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, November 2, 1964.

2. J. L. Kerrebrock, Nonequilibrium Ionization Due to Electron Heating: I. Theory, AIAA J. 2, 1072-1080 (1964).

G. BONDING MECHANISM OF ALKALI-METAL ATOMS ADSORBED ON METAL SURFACES

1. Introduction

Several illuminating theoretical studies of surface and adsorption phenomena of metal vapors on dissimilar metal substrates have been published.[1-3] An interesting point is that each treatment proposes an entirely different physical mechanism for cesium adsorption on metals. Despite this seeming variance, all three theories are able to explain experimental data relevant to cesium adsorption. These data consist of experimentally measured thermionic electron emission from cesiated surfaces and energies of adsorption of the cesium.

At this point a brief summary of the salient features of each of these analyses is in order. Gyftopoulos and Levine[1] propose that cesium is chemisorbed on the surface thereby forming a partially ionic-partially covalent bond with the four substrate atoms upon which the cesium rests, thereby forming essentially a CsW_4 molecule lying on the surface. To a large extent, they neglect the effects of the other substrate atoms. Their results agree well with experiment.

Gomer and Swanson[2] attempted a semiquantitative study to obtain knowledge of the basic physical mechanism giving rise to energies of adsorption. They feel that as an alkali atom such as cesium (in which the outer shell electron-energy level lies above the Fermi level) is brought to a metal surface, this level becomes greatly broadened. This broadening, which is analogous to natural broadening in plasmas, stems from the fact that the atom, brought into an interacting state with the metal, has a finite lifetime as an atom, and thus an uncertainty in its energy according to the uncertainty principle. They feel that the 6s level of cesium is broadened so greatly (several electron volts) that the atom level overlaps the conduction band of the substrate. It is then possible for polar metallic bonds to form between adsorbate and substrate. This can occur only if there is significant overlap of the broadened atomic level and the conduction band of the substrate.

Rasor and Warner[3] feel that cesium adsorption can be treated as if distinct species of atom and ions exist on the surface. The ratio of atoms and ions is obtained through a statistical mechanical treatment. For cesium on refractory metals at temperatures of 1000 °K most of the adsorbates are ions. Those ion adsorbates are bonded to the surface by purely ionic bonds. They justify this by arguments that include the necessity for the atomic level to be unbroadened. Through rough and questionable calculation they conclude that the level is broadened much less than 0.2 ev and thus their model is valid.

In this report, the high points of a detailed investigation[4] into the nature of the cesium-substrate bond are presented. Some of the ambiguity and confusion resulting from the differences summarized above will be removed. Detailed mathematics will be
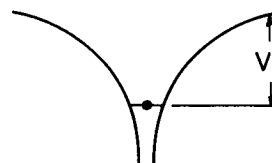
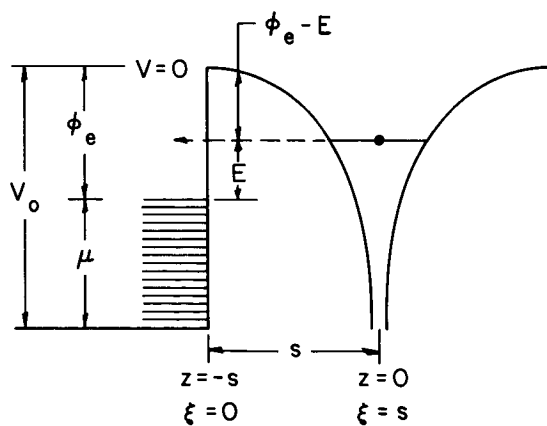Fig. XII-22. Sommerfeld metal.

Fig. XII-23. Isolated atom.

Fig. XII-24. Interacting metal and atom.

omitted, since it can be found elsewhere.[4] Only the basic physics and terminal results are reported herein.

2. Physical Model

The metal upon which cesium is adsorbed is considered to be a Sommerfeld metal. This is shown in Fig. XII-22. The isolated atom is shown in Fig. XII-23. In this figure, μ is the height of the zero temperature Fermi level; ϕ_e, the electron work function of the metal; and V_i, the unperturbed ionization potential of cesium.

As the atom at a distance s from the surface is allowed to interact with the metal, Fig. XII-24 may be considered. The dashed line indicates the transition of the electron from the atom to the metal; Z and ξ are simply coordinates; and E is defined by the figure in which $\phi_e - E$ is the perturbed ionization potential of the atom. Hagstrum[5] has shown both theoretically and experimentally that the conduction band of the metal retains all of its bulk properties out to the surface, so Fig. XII-24 is valid. Thus the idea of the CsW_4 molecule seems incorrect. Since the detailed nature of the surface barrier does not enter the ensuing calculations, the fact that a square well has been drawn in Fig. XII-24 is of no consequence.

3. Transition Process

As an atom is brought to a metal surface and is allowed to interact with the metal, the outer electron will want to make transitions into other energetically permissible quantum states such as are in the metal. If there is a perturbation coupling the two states, the transition probability can be determined. Calculations somewhat analogous to the work presented here have been made for Auger neutralization of ions by metals.[5-7]

The transition probability per unit time is given by the "Golden Rule" as

$$w = \frac{2\pi}{\hbar} \int \left| \left\langle \psi_m \left| H' \right| \psi_a \right\rangle \right|^2 \rho_k \, d\Omega \tag{1}$$

with ψ_m the final-state metal wave function; ψ_a the initial-state atom wave function; H' the perturbation mixing the two states; ρ_k the density of final states; and the integration performed over all directions of K in the final state.

For a Sommerfeld metal the following wave functions result:

$$\psi_m = \frac{1}{K_v L^{3/2}} \exp[i(K_{o1}x + K_{o2}y)] \left\{ (\underline{K}_{o3} + K_{o3}) \exp[i\underline{K}_{o3}\xi] + (\underline{K}_{o3} - K_{o3}) \exp[-i\underline{K}_{o3}\xi] \right\}$$

$$\text{for } \xi < 0 \tag{2}$$

$$\psi_m = \frac{1}{K_v L^{3/2}} \exp[i(K_{o1}x + K_{o2}y)] \, 2\underline{K}_{o3} \exp[i\underline{K}_{o3}\xi] \qquad \text{for } \xi > 0,$$

where

$$V_o = \frac{\hbar^2}{2m} K_o$$

$$K_o^2 = K_{o1}^2 + K_{o2}^2 + K_{o3}^2$$

$$\underline{K}_{o3}^2 = K_{o3}^2 + K_{v_o}^2$$

$$\underline{K}_o^2 = K_o^2 + K_{v_o}^2.$$

Here, subscripts 1, 2, 3 correspond to x, y, z directions, bars under K's indicate that the energy is measured from the bottom of the conduction band, and K_o and K_{o3} are both positive imaginary numbers.

For the atom, a hydrogen 2s wave function is fitted to the accurate Hartree-Fock 6s cesium wave function[8] to yield

$$\psi_a = \frac{a_1^{3/2}}{\sqrt{\pi}} (1-ar) \, e^{-ar}, \tag{3}$$

with $a = 0.755 \, \text{Å}^{-1}$, and $a_1 = 0.6 \, \text{Å}^{-1}$.

The perturbation is taken to be the potential of the ion core as seen by a metal electron. While this is the perturbation for the backward transition, not the transition from atom to metal, it is valid to use it for the perturbation of the atom-to-metal transition as has been pointed out by others.[9,10] Thus the perturbation is

$$H' = \frac{q^2}{r}. \tag{4}$$

Combining Eqs. 1 and 4, using the usual density of states function, and performing the necessary mathematical operations[4] the final result for the transition probability

$$w = \frac{4a_1^3 aq^2 \underline{K}s^3 \, e^{-2as}}{K_{v_o}^2 a \hbar} \left\{ 1 - \frac{4}{a^2 s^2} + \frac{8}{a^4 s^4} \right\}, \tag{5}$$

where $a_o = \hbar^2/mq^2$, and $\underline{K}$ is the wave number equivalent of the energy (measured from the bottom of the conduction band) of the electron involved in the transition. As can be seen, when the atom-metal separation goes to infinity (corresponding to no interaction), the transition probability goes to zero.

The effective lifetime τ of an atom before becoming an ion is given by the reciprocal of w, or

$$\tau = \frac{1}{w}.\tag{6}$$

This finite lifetime of the quantum state gives rise to a natural broadening of the atomic level, because of the uncertainty principle. The bandwidth is given by

$$\Gamma(s) \geq \hbar w(s) = \frac{\hbar}{\tau}.$$

4. Numerical Results

Using the reported values[11,12] of $\mu = 6.5$ ev, $E = 1.05$ ev, and $\phi_e = 4.6$ ev, for cesium on tungsten, Eq. 6 is evaluated as a function of atom distance from the surface. The lifetime τ and the bandwidth Γ are plotted as a function of s, in a hybrid but unambiguous manner in Fig. XII-25.
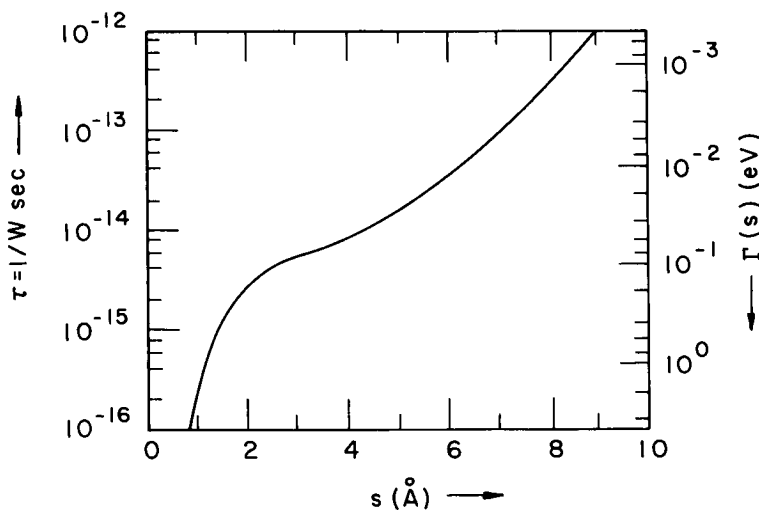

Fig. XII-25.   Atomic lifetime and bandwidth as a function of distance from the surface.

If the distance of an adsorbed particle from the surface is taken to be 2.65 Å, the usual value of the so-called atomic radius, then the lifetime of the atomic state on the surface is found to be $\sim 5 \times 10^{-15}$ second. The bandwidth is ~ 0.15 ev, a value lying between that required for validity of either Rasor's or Gomer's theory.

5. Discussion and Conclusions

Note that the mechanism for bonding of cesium upon a metal surface proposed by Rasor and by Gomer are inconsistent with the results of this analysis. Some of the fea-

tures of the Levine model, however, are in complete accord with the results presented here, and in fact require these results. This will now be explained.

If an atom is brought to a surface at zero temperature, three distinct types of bonds can be formed. A completely ionic bond occurs if there is no overlap of conduction band
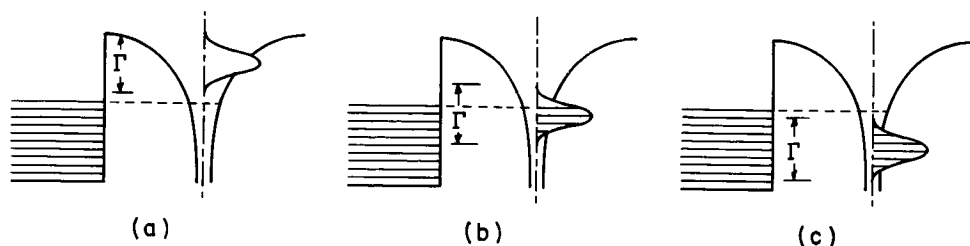

Fig. XII-26. (a) Ionic bonding. (b) Partially ionic-partially covalent bonding. (c) Polar metallic or covalent type of bonding.

and broadened atomic level as shown in Fig. XII-26a. This is the required situation for validity of Rasor's model. If the atom and metal have some overlap as shown in Fig. XII-26b, a partially ionic-partially covalent bond will form. Levine requires this situation, although it is not stated explicitly in his analysis. A word of caution on Levine's theory, though. He feels a partially ionic-partially covalent bond forms between adsorbate atoms and four substrate atoms. This is not really true, for a partially ionic-partially covalent bond is actually formed between the atomic electrons and the free conduction electron of the metal, not electrons associated with a particular substrate atom. Identifying a metal electron with a particular ion core is a meaningless concept. Figure XII-26c shows the conditions required for the adsorption mechanism proposed by Gomer.

As the temperature is raised above zero, some conduction electrons are found at levels overlapping the broadened level of the adsorbate as shown in Fig. XII-27, where
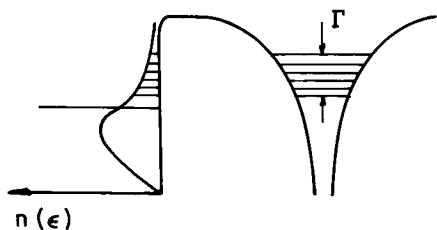

Fig. XII-27. Conduction-band electrons available for covalent bonding at $T > 0°$.

n (ε)

the distribution function is superposed on the metal-atom drawing. In this situation, the conditions being those that occur experimentally, the bond formed between adsorbate and metal would be partially ionic-partially covalent if the broadening of the atomic level is

by the amount calculated in this analysis. Thus it is understandable why the Levine model, despite the incorrect picture of a CsW_4 molecule, gives good results. It proposes the correct form of bond. The only reason that the ionic bond model gives good results is that it includes so many adjustable constants hidden under the label of physical constants.

A final check on the validity of the ionic-covalent bond will be made in the future. Levine has proposed that the energy of adsorption for the ionic-covalent bond is the geometric mean of the energy of adsorption for purely ionic and purely covalent bonding. The purely ionic bond energy of adsorption can be calculated from Rasor's model. We intend to compute the actual covalent bond energy by calculating the exchange energy for two electrons, one electron being in the metal at an energy within the energies of the broadened band, and the other within the atom. This is the theoretical purely covalent bond energy for an adsorbate and a metal. Following Levine's claim, based on the Pauling type of argument, we should obtain the theoretical atomic energy of adsorption. This should agree with experimentally determined values, and thus serve to prove that the bond formed between a cesium atom and a metal surface is partially ionic-partially covalent.

J. W. Gadzuk

References

1. E. P. Gyftopoulos and J. D. Levine, J. Appl. Phys. 33, 67 (1962).

2. R. Gomer and L. W. Swanson, J. Chem. Phys. 38, 1613 (1963).

3. N. S. Rasor and C. Warner, J. Appl. Phys. 35, 2589 (1964).

4. J. W. Gadzuk, Adsorption Physics of Alkali Metal Vapors on Metal Surfaces — Cesium-Tungsten, S. M. Thesis, Department of Mechanical Engineering, M. I. T., January 1965.

5. H. D. Hagstrum, Phys. Rev. 96, 336 (1954).

6. D. Sternberg, Ph. D. Thesis, Department of Physics, Columbia University, New York, 1957.

7. H. D. Hagstrum, Phys. Rev. 122, 83 (1961).

8. A. Russek, C. H. Sherman, and D. E. Flinchbough, Phys. Rev. 126, 573 (1962).

9. D. R. Bates, A. Fundaminsky, and M. S. W. Massey, Trans. Roy. Soc. (London) 243, 93 (1950).

10. D. R. Bates (ed.), Atomic and Molecular Processes (Academic Press, New York and London, 1962), p. 572.

11. M. F. Manning and M. I. Chodorow, Phys. Rev. 56, 787 (1939).

12. J. B. Taylor and I. Langmuir, Phys. Rev. 44, 423 (1933).

COMMUNICATION SCIENCES

AND

ENGINEERING

A. A SPECIAL CLASS OF QUANTIZER-INPUT SIGNALS

In a previous study,[1] the quantizer structure has been restricted in one case to have constrained transition values, and to have constrained representation values. (For nomenclature refer to Fig. XIII-1.) In each of these two cases sufficient conditions have been found on the error-weighting function that the quantization error surface had a
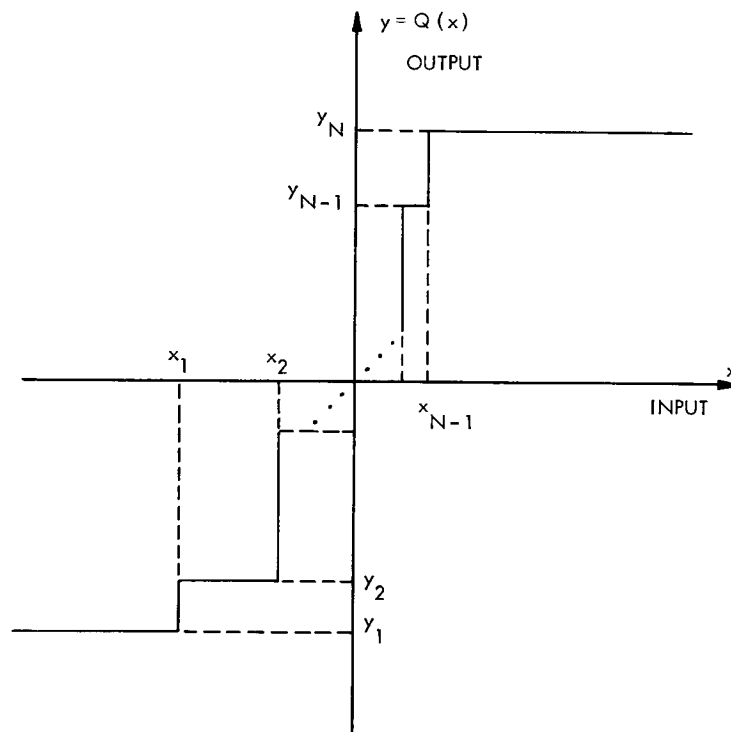

Fig. XIII-1. Input-output characteristic of the N-level quantizer. The x_i are the transition values and the y_i are the representation values.

single relative extremum, a relative minimum. In a recent paper,[2] P. E. Fleischer has constrained the error-weighting function to be $g(e) = e^2$. He then derived a sufficient condition on the amplitude probability density of the quantizer-input signal so that the associated error surface will also have one relative extremum, a relative minimum. Fleischer's sufficient condition is expressed by the inequality

$$\frac{\partial^2}{\partial \xi^2} \{\ln[p_x(\xi)]\} < 0. \tag{1}$$

where $p_x(\xi)$, the amplitude probability density of the quantizer-input signal, is required to be continuous. In order to derive this condition, he determines a sufficient condition on the matrix of second partial derivatives of the quantization error so that all relative extrema will be relative minima. This condition is equivalent to requiring that the matrix of second partials be positive definite. In determining the condition for which this matrix is positive definite Fleischer used the row-sum condition.[3] The condition that is obtained guarantees that the error surface will have only one relative extremum and that it will be a relative minimum.

In a typical practical problem involving experimental data $p_x(\xi)$ will be specified numerically, rather than by an algebraic expression. Since numerical differentiation cannot be accurately accomplished, this particular form of Fleischer's condition cannot be used to determine whether or not $p_x(\xi)$ has a single relative extremum. Our primary purpose here is to transform (1) into an equivalent condition that can be used to determine whether or not a specific numerically specified amplitude density satisfies Fleischer's condition.

We begin by recalling the definition of a strictly convex function.

A function $f(x)$ is strictly convex if and only if

$$f[\alpha a + (1-\alpha)b] < \alpha f(a) + (1-\alpha)f(b), \tag{2}$$

for all $b > a$ and all α such that $0 < \alpha < 1$. It can be determined, for example, by graphical consideration of the implications of Eq. 2, that strictly convex functions also satisfy the inequality

$$\frac{\partial^2}{\partial x^2}[f(x)] > 0. \tag{3}$$

Therefore, comparing Eqs. 1 and 3, we see that Fleischer's condition is equivalent to requiring that the function

$$\psi(\xi) = -\ln[p_x(\xi)] \tag{4}$$

be strictly convex. Observing that the strictly convex criteria (2), may be alternatively written

$$e^{-\psi[\alpha a+(1-\alpha)b]} > e^{-[\alpha\psi(a)+(1-\alpha)\psi(b)]} \tag{5}$$

for all b > a and for all α so that $0 < \alpha < 1$, we can write for (4)

$$p_x(\xi) = e^{-\psi(\xi)} \tag{6}$$

and, by direct substitution of (6) into (5), have

$$p_x[\alpha a+(1-\alpha)b] > [p_x(a)]^{\alpha}[p_x(b)]^{(1-\alpha)}. \tag{7}$$

If this inequality is satisfied for all b > a and all α so that $0 < \alpha < 1$, then Fleischer's condition is satisfied.

An examination of Eq. 7 indicates several properties of the amplitude probability densities which satisfy this condition. First, if we consider the case for which $p_x(a) = p_x(b)$, we find that (7) can be written equivalently as
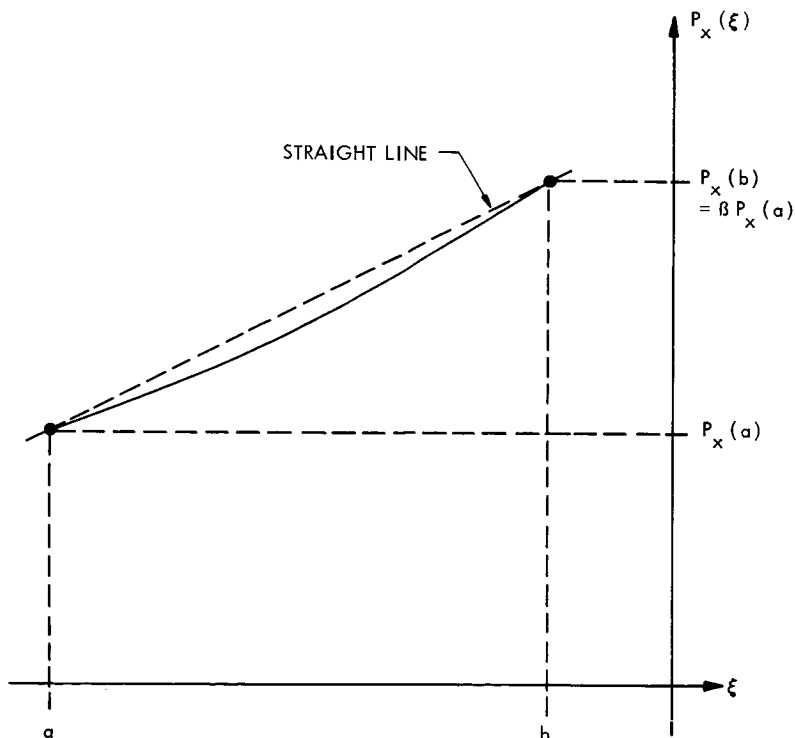

Fig. XIII-2. Illustration of Eq. 7 for $p_x(b) = \beta p_x(a)$. (The figure is drawn for $\beta = 2$.)

$$p_x[\alpha a + (1-\alpha)b] > p_x(a) = p_x(b)$$

or

$$p_x(\xi) > p_x(a) = p_x(b) \tag{8}$$

for $a < \xi < b$. This implies that the $p_x(\xi)$ satisfying this condition must have only one relative extremum and that this relative extremum is a relative maximum. Second, if we consider the case in which

$$p_x(b) = \beta p_x(a),$$

Eq. 7 becomes

$$p_x(\xi) > \beta^{(1-\alpha)} p_x(a) \tag{9}$$

for $a < \xi < b$. From a graphical examination of this condition (Fig. XIII-2), we see that the $p_x(\xi)$ that satisfy Fleischer's condition possess a type of mild convexity property.

In conclusion, we note that Eq. 7, the condition equivalent to Fleischer's condition, is one that can easily be utilized, for example, by means of a digital computer search program, to determine whether or not a given numerically obtained amplitude probability density satisfies Eq. 1.

J. D. Bruce

References

1. J. D. Bruce, An Investigation of Optimum Quantization, Sc. D. Thesis, Department of Electrical Engineering, M. I. T., June 1964; a report based on this thesis will appear as Technical Report 429, Research Laboratory of Electronics, M. I. T.

2. P. E. Fleischer, Sufficient Conditions for Achieving Minimum Distortion in a Quantizer, 1964 IEEE International Convention Record, Part 1, pp. 104-111.

3. R. Bellman, Introduction to Matrix Analysis (McGraw-Hill Book Company, Inc., New York, N. Y., 1960), pp. 294-295.

B. OPTIMUM HOMOMORPHIC FILTERS

1. Introduction

In this report optimum homomorphic filters will be discussed. The problem can be stated as follows: two signals $s_1(t)$ and $s_2(t)$ are combined according to some rule, denoted o. We wish to determine the optimum homomorphic system from the class having o as both the input and output operations such that the error between the output and the desired output is minimized. Hence, we wish to select a nonlinear system from a specified class of nonlinear systems which is optimum according to some error criterion.

The error criterion that will be used depends on the class of homomorphic systems under consideration. Specifically, the error criterion to be associated with any particular filtering problem will be restricted to be the norm of the error vector in the vector space of system outputs. This will guarantee that the error criterion satisfies the properties that we associate with error criteria such as mean-square error or integral-square error. Since the norm that we associate with a particular vector space is not unique, this restriction still permits flexibility in the specific error criterion that we select and still affords analytical convenience.

Our approach is to consider the meaning of the error criterion on the output space in the light of the cascade representation for homomorphic systems. We shall show that the characteristic system to associate with the class can be chosen in such a way that minimization of the error is equivalent to minimization of mean-square error or integral-square error (the choice depending on whether the system inputs are continuing or aperiodic) at the output of the linear portion of the cascade representation. When the characteristic systems have been selected, the choice of the optimum system reduces to a choice of the linear system in the cascade representation.

2. Linear Filtering Problem

The linear filtering problem, considered from the vector-space point of view, consists in choosing a linear operator on the space, which will separate a given vector from a linear combination of vectors in the space. When the vectors represent time functions, with vector addition as the algebraic addition of the functions, we usually impose additional constraints such as realizability and time invariance of the system represented by the linear operator. The primary aspect of the vector-space interpretation of filtering is that it implies the determination of a linear transformation for which the range is a subspace of the domain. In the general case, for which the linear transformation corresponds to a homomorphic system, the output space is a subspace of the input space. On this basis, it is clear that the classes of homomorphic systems which are of interest in this discussion are those for which the input and output operations are the same.

Consider a vector space V. Let v_1, v_2, ..., v_n be n vectors in V, and let v be a linear combination of these n vectors, that is,

$$v = \sum_{k=1}^{n} a_k v_k.$$

A necessary and sufficient condition so that there exists a linear transformation T_j on V such that

$$T_j(v) = a_j v_j$$

for all j, is that the set of vectors v_1, v_2, ..., v_n be linearly independent.[1] When restated in terms of linear systems, this means that there is always a linear system that will filter an input from a linear combination of inputs, provided only that the inputs that have been combined are linearly independent. This can also be interpreted to state that if the vectors v_1, v_2, ..., v_k are chosen from independent subspaces of V, then each of the vectors in the linear combination can be separated from the vector v. The implications of this with respect to linear multiplexing have been discussed by Zadeh and Miller.[2] The existence of the linear transformation T_j does not, of course, guarantee realizability or time invariance of the associated linear system.

When signal separation cannot be performed exactly because the vectors do not lie in independent subspaces or there are additional constraints to impose on the transformations, then an error criterion must be selected on the basis of which the optimum transformation can be chosen. When an inner product is defined on the vector space, this error criterion is generally selected to be the length, or norm, of the error vector. This choice is based on two considerations: (i) it is meaningful as an error criterion; and (ii) it is analytically convenient.

The error vector is defined as the vector sum of the desired output and the inverse of the actual output. That is,

$$v_e = v_d + (-v_o),$$

where v_e is the error vector, v_d is the vector representing the desired output, and v_o is the vector representing the actual output. The vector $(-v_o)$ represents the inverse of v_o. For example, if $s_d(t)$ is the desired output, $s_o(t)$ is the actual output and vector addition is the product of the time functions, then the time function corresponding to the error vector will be

$$e(t) = s_d(t)/s_o(t).$$

When an inner product is associated with the vector space, the norm of a vector is taken as the positive square root of the inner product of the vector with itself. The usefulness of the norm of the error vector as an error criterion arises directly from the algebraic properties of the norm. Specifically, let (v_1, v_2) denote the inner product of a vector v_1 with a vector v_2. Then for any vectors v_1, v_2, and v_3 in V and for any scalar c in the field associated with V, the properties of an inner product require that

(1) $(v_1, v_2) = \overline{(v_2, v_1)}$.

Here, the bar denotes conjugation.

(2) $(v_1 + v_2, v_3) = (v_1, v_3) + (v_2, v_3)$

(3) $(cv_1, v_2) = c(v_1, v_2)$

(4) $(v_1, v_1) > 0$ if and only if v_1 is not the zero vector.

The norm of a vector v denoted $\|v\|$ in an inner product space is given by

$$\|v\|^2 = (v, v).$$

It follows from properties (1), (2), (3), and (4) that

(1') $\|v\| > 0$ if and only if v is not the zero vector

(2') $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$

(3') $\|cv_1\| = |c| \|v_1\|$.

From property (1') it is clear that any norm for the error vector satisfies the fundamental restriction that we would like to impose on an error criterion. Specifically, it guarantees that positive and negative errors cannot cancel, since $\|v_e\| = 0$ implies that the actual output and desired output are equal.

3. Optimum Homomorphic Filters

The cascade representation for the class of homomorphic systems having the operation o as both input and output operations is shown in Fig. XIII-3. The inputs are restricted to constitute a Hilbert space with an orthonormal basis, under some inner product, and the set of outputs to constitute a subspace of the vector space of inputs. The cascade representation is derived by mapping the input space linearly in a one-to-one manner onto a space in which vector addition corresponds to algebraic addition. A linear operation is then performed on this space and the resulting outputs are mapped onto the space, which is a subspace of the inputs, by means of the inverse of the system wish to show first that if L is chosen on the basis of any error criterion that is a its output space, then the over-all system will be optimum under some error that is a norm on the output space of the over-all system. This conclusion is eorems I and II which follow.

1: Let V and W denote two inner product spaces, and T denote an r transformation between them. The inner product of two vectors v_i and written as (v_i, v_j), and the inner product of two vectors w_i and w_j in W the inner product $[w_i, w_j]$ in W can be taken as

$$\left(T^{-1}(w_i), T^{-1}(w_j) \right)$$

(1)

$$v_i = T^{-1}(w_i)$$

and

$$v_j = T^{-1}(w_j).$$

PROOF: We need only show that the inner product on W satisfies properties (1) as required of an inner product.

Property 1: $[w_1, w_2] = (v_1, v_2)$

$$[w_2, w_1] = (v_2, v_1).$$

But (v_1, v_2) is an inner product of V; hence

$$(v_1, v_2) = \overline{(v_2, v_1)}.$$

Therefore

$$[w_1, w_2] = \overline{[w_2, w_1]}.$$

Property 2: $[w_1+w_2, w_3] = (v_1+v_2, v_3),$

since T^{-1} is linear, and therefore

$$T^{-1}(w_1+w_2) = T^{-1}(w_1) + T^{-1}(w_2).$$

But $(v_1+v_2, v_3) = (v_1, v_3) + (v_2, v_3).$

Also, $[w_1, w_3] + [w_2, w_3] = (v_1, v_3) + (v_2, v_3);$ hence

$$[w_1+w_2, w_3] = [w_1, w_3] + [w_2, w_3].$$

Property 3: $[cw_1, w_2] = \left(T^{-1}(cw_1), T^{-1}(w_2) \right)$

$$= \left(cT^{-1}(w_1), T^{-1}(w_2) \right)$$

$$= (cv_1, v_2)$$

$$= c(v_1, v_2).$$

Therefore

$$[cw_1, w_2] = c(v_1, v_2)$$

or

$$[cw_1, w_2] = c[w_1, w_2].$$

Property 4: Since T is invertible and linear,

$$T^{-1}(w) = 0 \quad \text{if and only if } w = 0.$$

By definition of the inner product in W,

$$[w, w] = (T^{-1}(w), T^{-1}(w)).$$

But $(T^{-1}(w), T^{-1}(w)) = 0$ if and only if $T^{-1}(w) = 0$, since the inner product in V satisfies the required properties. Consequently,

$$[w, w] = 0 \quad \text{if and only if } T^{-1}(w) = 0$$

which in turn requires that

$$[w, w] = 0 \quad \text{if and only if } w = 0.$$

THEOREM 2: Let V and W denote two Hilbert spaces with orthonormal bases. Let (v_a, v_b) denote the inner product of any two vectors v_a and v_b in V, and $[w_c, w_d]$ denote the inner product of any two vectors w_c and w_d in W. Then, an invertible linear transformation T can always be defined which maps V to W in such a way that

$$[w_a, w_b] = \left(T^{-1}(w_a), T^{-1}(w_b) \right)$$

$$= (v_a, v_b)$$

for any vectors w_a and w_b in W.

PROOF: Let v_1, v_2, ... denote an orthonormal basis in V, and w_1, w_2, ... denote an orthonormal basis in W. Consider any vector v in V which, in terms of the basis in V, is

$$v = \sum_{k=1}^{\infty} (v, v_k) \, v_k.$$

Then the transformation T will be defined as

$$T(v) = \sum_{k=1}^{\infty} (v, v_k) \, w_k.$$

It has been shown that this transformation is linear and invertible.[3] The proof is omitted here to avoid a discussion of the convergence of the summation in Eq. 2. Now, consider any two vectors v_a and v_b in V, and let $T(v_a) = w_a$ and $T(v_b) = w_b$. The vector v_a is expressible as

$$v_a = \sum_{k=1}^{\infty} (v_a, v_k)\, v_k$$

and the vector v_b is expressible as

$$v_b = \sum_{k=1}^{\infty} (v_b, v_k)\, v_k.$$

Hence, $T(v_a) = w_a = \sum_{k=1}^{\infty} (v_a, v_k)\, w_k$ and $T(v_b) = w_b = \sum_{k=1}^{\infty} (v_b, v_k)\, w_k.$ The inner product of w_a and w_b is, then,

$$[w_a, w_b] = \left(\sum_{k=1}^{\infty} (v_a, v_k)\, w_k, \ \sum_{r=1}^{\infty} (v_b, v_r)\, w_r \right)$$

$$= \sum_{k=1}^{\infty} \sum_{r=1}^{\infty} (v_a, v_k)(v_r, v_b)[w_k, w_r].$$

But the set $w_1, w_2, \ldots$ is orthonormal in W; and hence

$$[w_k, w_r] = \begin{cases} 1 & k = r \\ 0 & k \neq r \end{cases}$$

Therefore

$$[w_a, w_b] = \sum_{k=1}^{\infty} (v_a, v_k)(v_k, v_b).$$

Similarly, the inner product of v_a and v_b is

$$(v_a, v_b) = \left(\sum_{k=1}^{\infty} (v_a, v_k)\, v_k, \ \sum_{r=1}^{\infty} (v_b, v_r)\, v_r \right)$$

$$= \sum_{k=1}^{\infty} \sum_{r=1}^{\infty} (v_a, v_k)(v_r, v_b)(v_k, v_r)$$

$$= \sum_{k=1}^{\infty} (v_a, v_k)(v_k, v_b)$$

since the set $v_1, v_2, \ldots$ is orthonormal in V. Thus

$$[w_a, w_b] = (v_a, v_b) = \sum_{k=1}^{\infty} (v_a, v_k)(v_k, v_b).$$

Theorem 1 tells us that given a linear invertible transformation between two vector spaces, a norm can be defined on the output space in such a way that the norm of a vector is invariant under the transformation. Alternatively, if the norm has already been specified on both spaces, then from Theorem 2 we can find an invertible linear transformation between these spaces under which the norm of a vector is invariant. It remains to interpret these theorems within the framework of homomorphic systems.
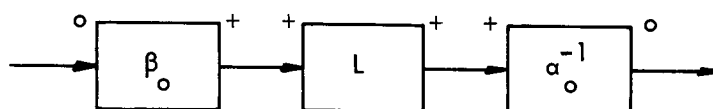

Fig. XIII-3. Cascade representation for the class of homomorphic systems having o as both the input and output operations.

The system represented in Fig. XIII-3 is a linear transformation mapping a vector space into itself. The output of the linear portion is a vector space V under addition (that is, with vector addition as the algebraic sum of the time functions) and the output of the over-all system is a vector space W under the operation o. The system a_o^{-1} is an invertible continuous linear transformation between these spaces. Let the norm of a vector v in V be rewritten as $\|v\|_V$, where restrictions are not placed on the norm other than those required by the definition of a norm. Similarly, let the norm of a vector w in W be written as $\|w\|_W$ defined as

$$\|w\|_W = \|a_o(w)\|_V \tag{3}$$

or equivalently,

$$\|v\|_V = \|a_o^{-1}(v)\|_W.$$

Suppose, now, that we wish to choose L such that the norm of the error vector in W is minimum. Let $f_d(t)$ be the desired output and $f_m(t)$ be the output for the optimum system, that is,

$$\|f_d \circ f^{-1}\|_W > \|f_d \circ f_m^{-1}\|_W \quad \text{for any } f \text{ in } W \text{ not equal to } f_m.$$

But $\|f_d \circ f^{-1}\|_W = \|a_o^{-1}(f_d) \circ a_o^{-1}(f^{-1})\|_V$. Hence

$$\|a_o^{-1}(f_d) \circ a_o^{-1}(f^{-1})\|_V > \|a_o^{-1}(f_d) \circ a_o^{-1}(f_m^{-1})\|_V$$

This can be rewritten in a more convenient form by using the fact that a linear transformation between vector spaces maps inverses to inverses, which requires that

$$\| a_o^{-1}(f_d) - a_o^{-1}(f) \|_V > \| a_o^{-1}(f_d) - a_o^{-1}(f_m) \|_V$$

for any f in W that is not equal to f_m. Hence, if the over-all system is optimized with respect to the norm defined by Eq. 3, then the linear portion of the cascade representation must be optimum, according to the norm defined in V, when the desired output of the linear portion is taken to be $a_o(f_d)$. Hence the problem of optimizing a homomorphic system within a specified class reduces to the optimization of the linear portion of the cascade representation relative to the desired output transformed by a_o. If the error criterion in W is specified, then the optimization of the linear system L must be carried out according to the error criterion specified by Eq. 3. If the error criterion in V is specified, then when the linear system L is optimum, the over-all system is optimum according to the error criterion dictated by Eq. 3.

As a result of Theorem 2, an alternative is offered. If an error criterion is specified in W, and we wish to optimize L according to a given error criterion in V, then the system a_o^{-1} can be constructed according to Theorem 2. Thus, if the error criterion is unspecified in W, we may make any convenient choice for the system a_o^{-1} and optimize L according to any convenient error criterion. If the error criterion is specified in W but unspecified in V, then a choice of a_o^{-1} will constrain the error criterion in V. Finally, if the error criterion is specified in both V and W, then when L is optimized the choice for a_o^{-1} is constrained if the over-all system is to be optimum. Thus far, we have been concerned with the choice of L and a_o^{-1} in the cascade representation. The question naturally arises as to how the optimization is affected by the system β_o. On the basis of the following theorem, it will be concluded that any choice for β_o can be made. The optimization of L will effectively compensate for the choice of β_o.

THEOREM 3: Let V and W be two vector spaces, and let β_a and β_b represent any two invertible homomorphic transformations from V onto W. Then there always exists a unique linear transformation L on W such that

$$\beta_a = L\beta_b,$$

where $L\beta_b$ represents the cascade of the transformations β_b and L. In other words, for every vector v in V, there exists L such that

$$\beta_a(v) = L[\beta_b(v)]. \tag{4}$$

PROOF: Consider the transformation L on W defined as

$$L(w) = \beta_a\left[\beta_b^{-1}(w)\right]. \tag{5}$$

The transformations β_a and β_b^{-1} are homomorphic, that is,

$$\beta_b^{-1}(w_1+w_2) = \beta_b^{-1}(w_1) + \beta_b^{-1}(w_2),$$

$$\beta_b^{-1}(cw) = c\beta_b^{-1}(w),$$

$$\beta_a(v_1+v_2) = \beta_a(v_1) + \beta_a(v_2)$$

and

$$\beta_a(cv) = c\beta_a(v).$$

We must show, first, that L is linear, that is,

$$L(w_1+w_2) = L(w_1) + L(w_2)$$

and

$$L(cw) = cL(w).$$

But

$$L(w_1+w_2) = \beta_a\left[\beta_b^{-1}(w_1+w_2)\right]$$

$$= \beta_a\left[\beta_b^{-1}(w_1) + \beta_b^{-1}(w_2)\right]$$

$$= \beta_a\left[\beta_b^{-1}(w_1)\right] + \beta_a\left[\beta_b^{-1}(w_2)\right]$$

and

$$\beta_a\left[\beta_b^{-1}(w_1)\right] = L(w_1)$$

$$\beta_a\left[\beta_b^{-1}(w_2)\right] = L(w_2).$$

Therefore $L(w_1+w_2) = L(w_1) + L(w_2)$. Similarly,

$$L(cw) = \beta_a\left[\beta_b^{-1}(cw)\right]$$

$$= \beta_a\left[c\beta_b^{-1}(w)\right]$$

$$= c\beta_a\left[\beta_b^{-1}(w)\right].$$

Hence $L(cw) = cL(w)$, and consequently L is linear. But L, defined in Eq. 5, provides the condition required by Eq. 4, for substituting Eq. 4 in Eq. 5, we obtain

$$\beta_a(v) = \beta_a\left[\beta_b^{-1}(\beta_b(v))\right]$$

$$= \beta_a(v).$$

It remains only to show that L as defined by Eq. 4 is unique. This will be done by showing that Eq. 4 implies Eq. 5.

Consider any transformation L having the property required by Eq. 4. If w is any vector in W, then there exists a vector v in V such that

$$v = \beta_b^{-1}(w),$$

since β_b is invertible. Substituting in Eq. 4, we have

$$\beta_a\left[\beta_b^{-1}(w)\right] = L\left[\beta_b\left(\beta_b^{-1}(w)\right)\right]$$

$$= L(w)$$

which is the transformation defined by Eq. 5.

Now, suppose that in the cascade representation of Fig. XIII-3, we have made an arbitrary choice for the system β_o. The error criterion used to optimize L, and the error criterion used to optimize the over-all system have been shown to determine the system α_o. Let L_{opt} denote the optimum choice for L. By virtue of Theorem 3, if the error could be further reduced by varying the system β_o, then the error could be reduced instead by cascading L_{opt} with an invertible linear system, which violates the assumption that L_{opt} is the optimum choice for L.

4. Example

In order to relate the preceding discussion to a specific example, consider the optimization of the systems in the class having multiplication as both input and output operations. This corresponds to the determination of an optimum filter for separating signals that have been multiplied, as is the case, for example, in transmission over a time-variant channel. It should be emphasized at the outset that for this example, the formalism that we have just completed does not seem justified by the intuitively reasonable result. The formalism indicates, however, that the approach is generally applicable to any class of homomorphic systems.
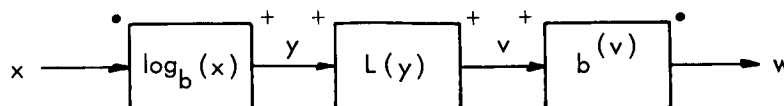
Fig. XIII-4. Cascade representation for the class of homomorphic systems having multiplication as both the input and output operations.

In the cascade representation for the class under consideration, the system a_o can be chosen as a logarithmic amplifier relative to any base b as shown in Fig. XIII-4, if the system inputs and outputs are restricted to be positive. Let us restrict the system L to be a linear, time-invariant, realizable system. If the input time functions for the over-all system are random, then a mean-square-error criterion is an analytically convenient choice for optimization of the linear system. Hence, if $f_d(t)$ is the desired output of the over-all system, then on the basis of the preceding discussion the system L is to be so chosen that the error E given by

$$E = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} [\log_b(f_d) - \log_b(f_o)]^2 \, dt \tag{6}$$

is minimum. If the input to the homomorphic system is $f_i(t)$, then

$$\log_{\beta_1}(f_o) = L[\log_{\beta_1}(f_i)].$$

The solution for the optimum system L is specified by the Weiner-Hopf equation.

The error E is also the error at the output of the exponential amplifier. If we change the error criterion in W but wish to maintain a mean-square-error criterion in V, then the characteristic systems a_o and β_o must be changed.

5. Conclusion

Homomorphic systems can be used for the separation of signals that have been non-additively combined in much the same way as linear systems are used for the separation of signals that have been linearly combined. The manner in which the signals are combined determines the class of homomorphic systems from which the optimum system is selected.

In determining optimum systems, the error criterion is generally based on consideration of its validity as a measure of error, and the convenience that it offers in carrying out the optimization. Since homomorphic systems are represented by linear transformations on vector spaces, the error criterion is restricted in this study to be describable as the norm of the error vector. This restriction does not completely specify the error criterion.

On the basis of the cascade representation for homomorphic systems, we have shown that the optimization within a class of homomorphic systems can be reformulated in terms of the optimization of the linear portion of the cascade representation. If the linear system is optimized according to some error criterion, then the over-all system will also be optimum according to some error criterion. In this case the error criterion under which the system is optimum will depend on the choice of the characteristic system used in the cascade representation to map the outputs of the linear system to the outputs of the over-all system. In particular, a mean-square-error criterion (or

integral-square error if the inputs to the linear system are aperiodic) can be used to determine the linear portion of the optimum filter. Alternatively, the error criteria for both the linear system and the over-all system can be selected independently. When this is done, the choice for the characteristic system is restricted.

If a mean-square or integral-square error criterion is used to optimize the linear system and this system is restricted to be time-invariant, then the optimization can be carried out by the methods developed by Wiener in his theory of optimum linear filtering. This does not guarantee that the total system will be time-invariant. To insure this, the characteristic systems must also be chosen to be time-invariant.

The ideas discussed in this paper may be potentially useful in the study of time-variant communication channels, for which the transmitted signal has effectively been multiplied by a noise waveform, and in the study of multipath channels in which signal and noise have been convolved. It is hoped that through this study further work and interest in this approach will be stimulated.

A. V. Oppenheim

References

1. K. Hoffman and R. Kunze, Linear Algebra (Prentice Hall, Inc., Englewood Cliffs, N. J., 1961), Chap. 3.

2. L. A. Zadeh and K. S. Miller, Fundamental aspects of linear multiplexing, Proc. IRE 40, 1091-1097 (1952).

3. A. V. Oppenheim, Superposition in a Class of Nonlinear Systems, Sc. D. Thesis, Department of Electrical Engineering, M. I. T., 1964, p. 47.

C. A SINGULAR MAXIMIZATION PROBLEM WITH APPLICATIONS TO FILTERING

In this report we generalize a previous result[1] on a problem of extrema under constraints and discuss its application to filtering and estimation.

The mathematical problem is the following: We would like to find the probability density p(x) that maximizes or minimizes the functional

$$L = \int L(x)\, p(x)\, dx \quad \text{under M constraints}$$

$$G = \int G(x)\, p(x)\, dx$$

$$F = \int F(x)\, p(x)\, dx$$

$$\vdots$$

$$1 = \int p(x)\, dx.$$

We require the range of the random variable x to be bounded to some finite interval [A, B].

We shall show that if there exists a p(x) satisfying all of the constraints, then the probability density that maximizes or minimizes L under M constraints is made up of M impulses, at most.

PROOF: Assume that there exists $p_0(x)$ such that all M constraints are satisfied and which results in some value L_0 of the functional L. In Fig. XIII-5 we define the regions R_1, R_2 ... R_{M+1} by arbitrary cuts of the base $[A, B]$.
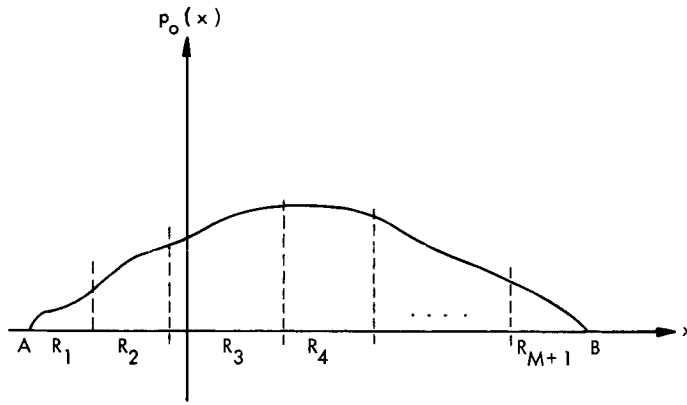

Fig. XIII-5. Probability density $p_0(x)$ that satisfies all constraints.

We shall show that there exists a probability density $p_1(x)$ that is nonzero on only M of the M+1 regions, which satisfies the M constraints and gives $L_1 \leq L_0$. Similarly, $p_2(x)$, which is nonzero on M regions at most, satisfies the M constraints, and yields $L_2 \geq L_0$. It then becomes clear that by successive applications of this result, the base of the probability density p(x) is reduced, at most, to M arbitrary small regions. Therefore the probability density p(x) that maximizes or minimizes L under M constraints will consist, at most, of M impulses.

To show that $p_1(x)$ and $p_2(x)$ with the properties claimed do exist, we define

$$m_1 = \int_{R_1} p_0(x)\, dx$$

$$F_1 = \frac{1}{m_1} \int_{R_1} F(x)\, p_0(x)\, dx$$

$$G_1 = \frac{1}{m_1} \int_{R_1} G(x)\, p_0(x)\, dx$$

.
.
.

$$L_{o1} = \frac{1}{m_1} \int_{R_1} L(x) \, p_o(x) \, dx$$

$$m_2 = \int_{R_2} p_o(x) \, dx$$

$$F_2 = \frac{1}{m_2} \int_{R_2} F(x) \, p_o(x) \, dx$$

$$\vdots$$

$$L_{o2} = \frac{1}{m_2} \int_{R_2} L(x) \, p_o(x) \, dx$$

etc.

We have, therefore, the new set of equations

$$m_1 L_{o1} + m_2 L_{o2} + \ldots + m_{M+1} L_{oM+1} = L_o$$

$$m_1 G_1 + \ldots + m_{M+1} G_{M+1} = G$$

$$\vdots$$

$$m_1 + \ldots + m_{M+1} = 1$$

and since $p(x) \geq 0$, we have also

$$m_1, \, m_2 \, \ldots \, m_{M+1} \geq 0.$$

Now consider these equations in the M+1 dimensional space of the weights $\underline{m}$. The M+1 equations define a point in this space, and any M equations define a line. Consider the last M equations. They define a line C_1 which intersects the subspace $m_1 L_{o1} + m_2 L_{o2} + \ldots + m_{M+1} L_{oM+1} = L_o$ at one point. The situation is represented in three dimensions in Fig. XIII-6.

On the constraint line C_1 all points not in the subspace $L = L_o$ correspond to $L > L_o$ of $L < L_o$. The line C_1 will leave the subspace of all $m_i > 0$, (i=1...M+1) at two points, say, $m_k = 0$ $m_e = 0$. These two points are necessarily on opposite sides of the subspace $L = L_o$. Therefore, we have $L > L_o$ at one point, and $L < L_o$ at the other. Therefore, in the space $\underline{m}$, these two points satisfy all constraints and give smaller and larger values to the functional L. In terms of $p(x)$, these two points correspond to a reassignment of the weights in the M+1 regions, making one of the weights zero, which is our result.

It is of interest to discuss briefly the reason why M impulses are needed for the number of constraints used. Assume that we subdivide $p(x)$ in M regions instead of M+1; we have, then, M+1 equations specifying one point in M-dimensional space. The M
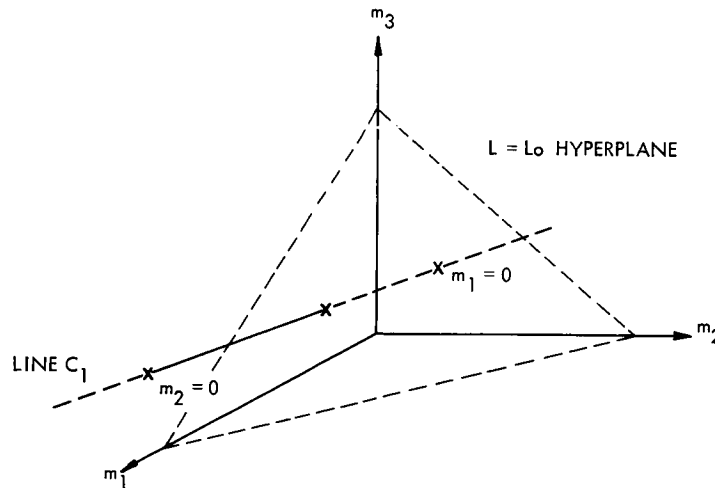
Fig. XIII-6. Discussion of the weight vector m in three dimensions.

constraints alone define one point in this M-dimensional space, the point we started from, and we cannot reassign weights in the M regions to increase or decrease L. Similarly, if we started with M+2 regions, then in the (M+2)-dimensional space, we satisfy M constraints by constraining $\underline{m}$ to a plane, and this allows us to make two of the weights equal to zero. The present result has been established previously[1] for M=2 by a more involved method, which did not lead simply to an extension.

GENERALIZATION: From the method of proof, it is clear that the result is not limited to a one-dimensional probability density p(x) and will hold for probability densities with any finite number of dimensions, as long as the range of $\underline{x} = (x_1, \ldots x_R)$ is bounded for all the variables. Using a vector notation, we have immediately the fact that is $p(\underline{x})$ satisfies M constraints of the form

$$\int G(\underline{x}) \, p(\underline{x}) \, \underline{dx} = G,$$

then the functional L, $L = \int L(\underline{x}) \, p(\underline{x}) \, \underline{dx}$ will be maximized or minimized for a probability density $p(\underline{x})$ which consists of M impulses at most.

Applications

1. Extrema of the Average Weighted Error for a Given Nonlinear Filter

Consider an input signal x(t), x(t) = m(t) + n(t) in which m(t) and n(t) are statistically independent processes. We desire to estimate the message m(t) at time $t = \{t_1 \ldots t_k\}$ from knowledge of the input with the same time constant. The noise statistics are known. Let

$$\hat{\underline{m}} = \{\hat{m}_1, \hat{m}_2, \ldots \hat{m}_k\}$$

be the estimates of $m(t_1)$, $m(t_2) \ldots$, and assume that the filter characteristic is known. We have

$$\hat{\underline{m}} = \underline{g}(\underline{x})$$

in which $\underline{g}(\underline{x})$ represents a set of k functions of the k input amplitudes. Let the error-weighting function be a scalar function of k variable $W(\underline{e})$ such as $W(\underline{e}) = e_1^2 + e_2^2 + \ldots + e_k^2$. Then the average weighted error can be written

$$\overline{W(\underline{e})} = \iint W[\underline{\lambda} - \underline{g}(\underline{\xi})]\, p_{\underline{m}, \underline{x}}(\underline{\lambda}, \underline{\xi})\, d\underline{\lambda}\, d\underline{\xi}.$$

But

$$p_{\underline{m}, \underline{x}}(\underline{\lambda}, \underline{\xi}) = p_{\underline{m}}(\underline{\lambda})\, p_{\underline{n}}(\underline{\xi} - \underline{\lambda}),$$

so that

$$\overline{W(\underline{e})} = \iint W[\underline{\lambda} - \underline{g}(\underline{\xi})]\, p_{\underline{n}}(\underline{\xi} - \underline{\lambda})\, p_{\underline{m}}(\underline{\lambda})\, d\lambda d\underline{\xi}.$$

If we let

$$L(\underline{\lambda}) \triangleq \int W[\underline{\lambda} - \underline{g}(\underline{\xi})]\, p_{\underline{n}}(\underline{\xi} - \underline{\lambda})\, d\underline{\xi}$$

in which $L(\underline{\lambda})$ is a known function of $\underline{\lambda}$, then

$$\overline{W(\underline{e})} = \int L(\underline{\lambda})\, p_{\underline{m}}(\underline{\lambda})\, d(\underline{\lambda}). \tag{1}$$

The problem considered is that of finding the specific message density, $p_{\underline{m}}(\underline{\lambda})$, that will maximize or minimize expression (1) under a set of message constraints. This is the mathematical problem discussed earlier; and if the message probability density $p(\underline{x})$ satisfies M constraints of the form $\int G(\underline{\lambda})\, p_{\underline{m}}(\underline{\lambda})\, d\lambda = G$, then $\overline{W(\underline{e})}$ will be maximizes or minimized if $p_{\underline{m}}(\underline{\lambda})$ consists of M impulses at most.

2. Minimum of the Average Weighted Error for the Optimum Nonlinear Filter

Since for any given nonlinear filter we have found that the minimum average weighted error corresponds to an impulsive message probability density, it is a simple matter to show that the same result holds if minimization is carried out among the filters, as well as among the message probability densities. Assume that $p_{\underline{m}o}(\underline{\lambda})$ and $\underline{g}_o(\underline{x})$ are the message probability density and the corresponding optimum nonlinear filter, and that the minimum average weighted error is $\overline{W_o(\underline{e})}$. For the filter $\underline{g}_o(\underline{x})$ considered as given

there exitst an impulsive message probability density $p_{m1}(\lambda)$ such that $\overline{W_1(e)} \leq \overline{W_0(e)}$. Also the optimum nonlinear filter $\underline{g}_1(\underline{x})$ associated with $p_{m1}(\lambda)$ will give $\overline{W_2(e)}$ such that

$$\overline{W_2(e)} \leq \overline{W_1(e)} \leq \overline{W_0(e)}. \tag{2}$$

Since, by hypothesis, $\overline{W_0(e)}$ is the lowest value achievable, we have necessarily equality in Eq. 2; therefore, the lowest average weighted error, $\overline{W_{min}(e)}$ is given by an impulsive message probability density $p_{m1}(\lambda)$.

Some examples for M=2 have been presented previously.[1] A more comprehensive discussion of these results will appear elsewhere.

V. R. Algazi

References

1. V. R. Algazi, A Study of the Performance of Linear and Nonlinear Filters, Technical Report 420, Research Laboratory of Electronics, M. I. T., June 22, 1964.

D. ON THE RELATION BETWEEN INTEGRAL AND DIFFERENTIAL CHARACTERIZATION OF NONLINEAR SYSTEMS

Consider the nonlinear system shown in Fig. XIII-7; N_1, N_2, and N_3 are linear systems and N_4 is a multiplier. The behavior of the system can be characterized by the set

$$\frac{dz(t)}{dt} + az(t) = x(t) \tag{1}$$

$$\frac{dw(t)}{dt} + bw(t) = x(t) \tag{2}$$

$$\frac{d^2y(t)}{dt^2} + d\frac{dy(t)}{dt} + ey(t) = cr(t) + \frac{dr(t)}{dt} \tag{3}$$

$$r(t) = w(t)\, z(t), \tag{4}$$

where $x(t)$ is the input, $y(t)$ is the output, and $w(t)$, $z(t)$, and $r(t)$ are the inputs and output of the multiplier, as shown in Fig. XIII-7. Equations 1-4 describe the behavior of N_1 through N_4. We assume that all initial conditions are zero.

We would like to find a differential equation relating $y(t)$ and $x(t)$; that is, we would like to eliminate $w(t)$, $z(t)$, and $r(t)$ in Eqs. 1-4. In order to do so, we shall extend the domain of definition from a line to a plane, and look along the 45° line in the plane.

Define $\hat{r}(t_1, t_2) = w(t_1)\, z(t_2)$, and $\hat{y}(t_1, t_2)$ so that $y(t)$ is $\hat{y}(t_1, t_2)\Big|_{t_1 = t_2 = t}$. Substitute t_1 for t in (2) and t_2 for t in (1). Then multiplication of (1) and (2) and use of the definition

of $\hat{r}(t_1, t_2)$ yields

$$\frac{\partial^2 \hat{r}(t_1, t_2)}{\partial t_2 \partial t_1} + a \frac{\partial \hat{r}(t_1, t_2)}{\partial t_2} + b \frac{\partial \hat{r}(t_1, t_2)}{\partial t_1} + ab\hat{r}(t_1, t_2) = x(t_1) \, x(t_2) \tag{5}$$

In order to express (3) in terms of $y(t_1, t_2)$ we must find an expression for $\frac{dy(t)}{dt}$ in terms of $\hat{y}(t_1, t_2)$. Since $\hat{y}(t, t) = y(t)$, the desired derivative will be the directional derivative
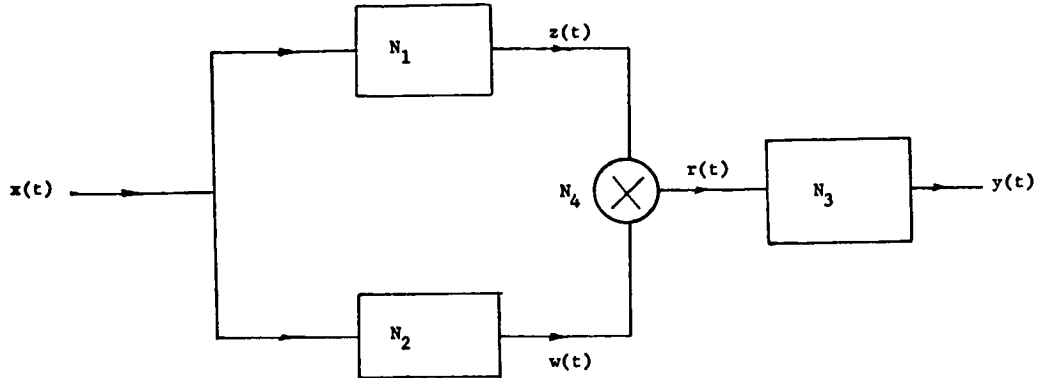

Fig. XIII-7. A simple nonlinear system.

of $\hat{y}(t_1, t_2)$ along the line $t_1 = t_2$, scaled by the factor $\sqrt{2}$ to obtain the proper rate of change. The directional derivative is given by the dot product of the gradient of $\hat{y}(t_1, t_2)$ with the unit vector in the direction of the 45° line. Hence we have the correspondence

$$\frac{dy(t)}{dt} \longleftrightarrow \sqrt{2} \left[\nabla \hat{y}(t_1, t_2) \cdot \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \right] = \frac{\partial \hat{y}(t_1, t_2)}{\partial t_1} + \frac{\partial \hat{y}(t_1, t_2)}{\partial t_2}. \tag{6}$$

Repeating this operation, we find the correspondence for the second derivative.

$$\frac{d^2 y(t)}{dt^2} \longleftrightarrow \frac{\partial^2 \hat{y}(t_1, t_2)}{\partial t_1^2} + 2 \frac{\partial^2 \hat{y}(t_1, t_2)}{\partial t_2 \partial t_1} + \frac{\partial^2 \hat{y}(t_1, t_2)}{\partial t_2^2}. \tag{7}$$

By using these results, (3) can be extended to

$$\frac{\partial^2 \hat{y}}{\partial t_1^2} + 2 \frac{\partial^2 \hat{y}}{\partial t_2 \partial t_1} + \frac{\partial^2 \hat{y}}{\partial t_2^2} + d \frac{\partial \hat{y}}{\partial t_1} + d \frac{\partial \hat{y}}{\partial t_2} + e\hat{y} = c\hat{r} + \frac{\partial \hat{r}}{\partial t_1} + \frac{\partial \hat{r}}{\partial t_2}. \tag{8}$$

We must now combine (8) and (5) to eliminate $\hat{r}(t_1, t_2)$. This may be accomplished by taking the partial of both sides of (5) with respect to t_1 to obtain (9), and with respect

to t_2 to obtain (10).

$$\frac{\partial^3 \hat{r}}{\partial t_2 \partial^2 t_1} + a \frac{\partial^2 \hat{r}}{\partial t_2 \partial t_1} + b \frac{\partial^2 \hat{r}}{\partial t_1^2} + ab \frac{\partial \hat{r}}{\partial t_1} = \frac{dx(t_1)}{dt_1} x(t_2) \tag{9}$$

$$\frac{\partial^3 \hat{r}}{\partial t_2^2 \partial t_1} + a \frac{\partial^2 \hat{r}}{\partial t_2^2} + b \frac{\partial^2 \hat{r}}{\partial t_2 \partial t_1} + ab \frac{\partial \hat{r}}{\partial t_2} = x(t_1) \frac{dx(t_2)}{dt_2}. \tag{10}$$

Also, we take the partial of (8) with respect to t_1 to obtain (11), with respect to t_2 to obtain (12), and with respect to t_1 and t_2 to obtain (13).

$$\frac{\partial^3 \hat{y}}{\partial t_1^3} + 2 \frac{\partial^3 \hat{y}}{\partial t_2 \partial t_1^2} + \frac{\partial^3 \hat{y}}{\partial t_2^2 \partial t_1} + d \frac{\partial^2 \hat{y}}{\partial t_1^2} + d \frac{\partial^2 \hat{y}}{\partial t_2 \partial t_1} + e \frac{\partial \hat{y}}{\partial t_1} = c \frac{\partial \hat{r}}{\partial t_1} + \frac{\partial^2 \hat{r}}{\partial t_1^2} + \frac{\partial^2 \hat{r}}{\partial t_2 \partial t_1} \tag{11}$$

$$\frac{\partial^3 \hat{y}}{\partial t_2 \partial t_1^2} + 2 \frac{\partial^3 \hat{y}}{\partial t_2^2 \partial t_1} + \frac{\partial^3 \hat{y}}{\partial t_2^3} + d \frac{\partial^2 \hat{y}}{\partial t_2 \partial t_1} + d \frac{\partial^2 \hat{y}}{\partial t_2^2} + e \frac{\partial \hat{y}}{\partial t_2} = c \frac{\partial \hat{r}}{\partial t_2} + \frac{\partial^2 \hat{r}}{\partial t_2 \partial t_1} + \frac{\partial^2 \hat{r}}{\partial t_2^2} \tag{12}$$

$$\frac{\partial^4 \hat{y}}{\partial t_2 \partial t_1^3} + 2 \frac{\partial^4 \hat{y}}{\partial t_2^2 \partial t_1^2} + \frac{\partial^4 \hat{y}}{\partial t_2^3 \partial t_1} + d \frac{\partial^3 \hat{y}}{\partial t_2 \partial t_1^2} + d \frac{\partial^3 \hat{y}}{\partial t_2^2 \partial t_1} + e \frac{\partial^2 \hat{y}}{\partial t_2 \partial t_1} = c \frac{\partial^2 \hat{r}}{\partial t_2 \partial t_1} + \frac{\partial^3 \hat{r}}{\partial t_2 \partial t_1^2} + \frac{\partial^3 \hat{r}}{\partial t_2^2 \partial t_1}. \tag{13}$$

Next, multiply through (5) by c, and add the resulting equation to (9) plus (10). Multiply through (12) by a, (11) by b, and (8) by ab, and add the sum of these new equations to (13). We can then write

$$cx(t_1) x(t_2) + x(t_1) \frac{dx(t_2)}{dt_2} + \frac{dx(t_1)}{dt_1} x(t_2) =$$

$$\frac{\partial^3 \hat{r}}{\partial t_2^2 \partial t_1} + \frac{\partial^3 \hat{r}}{\partial t_2 \partial t_1^2} + (a+b+c) \frac{\partial^2 \hat{r}}{\partial t_2 \partial t_1} + b \frac{\partial^2 \hat{r}}{\partial t_1^2} + a \frac{\partial^2 \hat{r}}{\partial t_2^2} + (ab+bc) \frac{\partial \hat{r}}{\partial t_1} + (ab+ac) \frac{\partial \hat{r}}{\partial t_2} + abc\hat{r} =$$

$$\frac{\partial^4 \hat{y}}{\partial t_2 \partial t_1^3} + 2 \frac{\partial^4 \hat{y}}{\partial t_2^2 \partial t_1^2} + \frac{\partial^4 \hat{y}}{\partial t_2^3 \partial t_1} + (a+2b+d) \frac{\partial^3 \hat{y}}{\partial t_2 \partial t_1^2} + (2a+b+d) \frac{\partial^3 \hat{y}}{\partial t_2^2 \partial t_1} + b \frac{\partial^3 \hat{y}}{\partial t_1^3}$$

$$+ a \frac{\partial^3 \hat{y}}{\partial t_2^3} + (ad+2ab+bd+e) \frac{\partial^2 \hat{y}}{\partial t_2 \partial t_1} + (ab+bd) \frac{\partial^2 \hat{y}}{\partial t_1^2} + (ab+ad) \frac{\partial^2 \hat{y}}{\partial t_2^2}$$

$$+ (abd+be) \frac{\partial \hat{y}}{\partial t_1} + (abd+ae) \frac{\partial \hat{y}}{\partial t_2} + abe\hat{y},$$

or

$$\frac{\partial^4 \hat{y}}{\partial t_2 \partial t_1^3} + 2\frac{\partial^4 \hat{y}}{\partial t_2^2 \partial t_1^2} + \frac{\partial^4 \hat{y}}{\partial t_2^3 \partial t_1} + (a+2b+d)\frac{\partial^3 \hat{y}}{\partial t_2 \partial t_1^2} + (2a+b+d)\frac{\partial^3 \hat{y}}{\partial t_2^2 \partial t_1} + b\frac{\partial^3 \hat{y}}{\partial t_1^3}$$

$$+ a\frac{\partial^3 \hat{y}}{\partial t_2^3} + (ad+2ab+bd+e)\frac{\partial^2 \hat{y}}{\partial t_2 \partial t_1} + (ab+bd)\frac{\partial^2 \hat{y}}{\partial t_1^2} + (ab+ad)\frac{\partial^2 \hat{y}}{\partial t_2^2}$$

$$+ (abd+be)\frac{\partial \hat{y}}{\partial t_1} + (abd+ae)\frac{\partial \hat{y}}{\partial t_2} + abe\hat{y} =$$

$$cx(t_1)\, x(t_2) + x(t_1)\frac{dx(t_2)}{dt_2} + \frac{dx(t_1)}{dt_1}\, x(t_2).$$

$$(14)$$

We have thus obtained a single differential equation relating $y(t)$ and $x(t)$. The equation is a linear partial differential equation with constant coefficients. This linear partial differential equation is particularly well suited to solution by means of the two-dimensional Laplace transform. Taking the transform of each side, we find

$$\left[s_1^3 s_2 + s_1^2 s_2^2 + s_1 s_2^3 + (a+2b+d)s_1^2 s_2 + (2a+b+d)s_1 s_2^2 + bs_1^3 + as_2^3 + (ad+2ab+bd+s)s_1 s_2 \right.$$

$$\left. + (ab+bd)s_1^2 + (ab+bd)s_2^2 + (abd+be)s_1 + (abd+ae)s_2 + abe \right] \hat{Y}(s_1 s_2) = (s_1+s_2+c)\, X(s_1)\, X(s_2)$$

Factoring the polynomials in this expression and solving for $\hat{Y}(s_1, s_2)$, we have

$$\hat{Y}(s_1, s_2) = \frac{s_1 + s_2 + c}{(s_1+b)(s_2+a)\left[(s_1+s_2)^2 + d(s_1+s_2) + e\right]}\, X(s_1)\, X(s_2). \qquad (15)$$

We now note that

$$H_2(s_1, s_2) \triangleq \frac{s_1 + s_2 + c}{(s_1+b)(s_2+a)\left[(s_1+s_2)^2 + d(s_1+s_2) + e\right]}$$

is the transform of the Volterra kernel, $h_2(\tau_1, \tau_2)$ of the system of Fig. XIII-7 when the system is characterized by the integral equation

$$y(t) = \iint h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2. \qquad (16)$$

In fact, taking the inverse transform of (15), we have

$$\hat{y}(t_1, t_2) = \iint h_2(\tau_1, \tau_2)\, x(t_1-\tau_1)\, x(t_2-\tau_2)\, d\tau_1 d\tau_2$$

from which (16) follows by setting $t_1=t_2=t$.

From this example the following observations may be made. Given a system of equations which is the dynamic description of a nonlinear system, we can. by extending the domain of definition from one dimension to two dimensions, find a single linear partial differential equation that also characterizes the system. That is, by extending from one dimension to two dimensions, a one-dimensional nonlinear problem has been converted into a two-dimensional linear problem.

Equations 1-4 and Eq. 14 describe the same situation. A system that is characterized by a single integral equation is equivalently described by a set of several ordinary differential equations and a nondifferential equation. A description by one nonlinear ordinary differential equation does not seem to be possible.

Whenever a system is characterized by an nth-degree Volterra kernel that has a rational transform, a linear partial differential equation with constant coefficients can be found relating the auxiliary output function $\hat{y}(t_1, \ldots, t_n)$ to the input function $x(t)$. If the kernel is of the class that can be realized exactly with a finite number of linear systems and multipliers,[1] then an equivalent description by a set of ordinary differential equations and nondifferential equations can be found.

Although the example and observations presented here have not yet led to the solution of any problems not easily handled by other methods, we feel that the viewpoint presented is unique and may lead to a deeper understanding of the properties of nonlinear systems.

A. M. Bush

References

1. A. M. Bush, Kernels Realizable Exactly with a Finite Number of Linear Systems and Multipliers, Quarterly Progress Report No. 76, Research Laboratory of Electronics, M. I. T., January 15, 1965, pp. 167-179.

E. USEFUL EXPRESSIONS FOR OPTIMUM LINEAR FILTERING IN WHITE NOISE

In this report, optimum linear filters for extracting a message from an additive white noise are considered. Results that have been derived may be listed as follows:

1.
$$H_{opt}(s) = 1 - \frac{N_o^{1/2}}{[S_a(-s^2) + N_o]^+} \tag{1}$$

2. $\overline{\epsilon^2_{min}}$, the minimum mean-square error, is given by

$$\overline{\epsilon^2_{min}} = N_o h_{opt}(0+) = N_o \lim_{s \to \infty} s H_{opt}(s) \tag{2}$$

$$\overline{\epsilon_{min}^2} = \frac{N_o}{2\pi} \int_{-\infty}^{\infty} \log \left[1 + \frac{S_a(\omega^2)}{N_o} \right] d\omega. \tag{3}$$

3. $H_{opt}(s)$ is minimum phase.

Equation 2 is especially convenient when the optimum filter is known. On the other hand, Eq. 3 is useful because it makes possible the determination of $\overline{\epsilon_{min}^2}$ without $H_{opt}(s)$.

Equations 1 and 3 have found wide application in the study of analog demodulation theory, their existence being crucial in studies by Develet,[2] Viterbi and Cahn,[3] and Van Trees and Boardman.[4]

Equations 1 and 3 have been derived previously by Yovits and Jackson,[1] and more recently by Viterbi and Cahn.[3] Here, all results are derived by use of a well-known expression for the optimum filter. No minimization arguments are required, as was previously the case.

1. Derivation of the Expression for $H_{opt}(s)$

The optimum linear filter is given by

$$H_{opt}(s) = \frac{1}{[S_a(-s^2) + N_o]^+} \left[\frac{S_a(-s^2)}{[S_a(-s^2) + N_o]} \right]_+, \tag{4}$$

where $S_a(\omega^2)$ and N_o are the power densities of the signal and noise, respectively.[5-7] $S_a(\omega^2)$ is taken to be a rational function of ω^2; the superscripts + and – indicate spectral factorization; the subscript + indicates the operation of taking the realizable part of a partial fraction expansion.

It is seen from Eq. 4 that

$$H_{opt}(s) = \frac{1}{[S_a(-s^2) + N_o]^+} \left[\frac{S_a(-s^2) + N_o}{[S_a(-s^2) + N_o]^-} - \frac{N_o}{[S_a(-s^2) + N_o]^-} \right]_+$$

$$= 1 - \frac{1}{[S_a(-s^2) + N_o]^+} \left[\frac{N_o}{[S_a(-s^2) + N_o]^-} \right]_+. \tag{5}$$

The term with the subscript + in Eq. 5 can only be a constant. Under the assumption that $\lim_{\omega \to \infty} S_a(\omega^2) = 0$, the constant is $N_o^{1/2}$. Equation 1 follows upon substitution of this constant in Eq. 5.

2. Derivation of $\overline{\epsilon_{min}^2}$ in Terms of the Optimum Filter

The minimum mean-square error is given by

$$\overline{\epsilon_{min}^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_a(\omega^2) \left| 1-H_{opt}(\omega) \right|^2 d\omega + \frac{1}{2\pi} \int_{-\infty}^{\infty} N_o \left| H_{opt}(\omega) \right|^2 d\omega. \tag{6}$$

From Eq. 1, it follows that

$$\left| 1-H_{opt}(\omega) \right|^2 = \frac{N_o}{S_a(\omega^2) + N_o} \tag{7}$$

and

$$\left| H_{opt}(\omega) \right|^2 = -\frac{N_o S_a(\omega^2)}{S_a(\omega^2) + N_o} + 2 \left| H_{opt}(\omega) \right| \cos \phi(\omega).$$

By using these two expressions in Eq. 6, $\overline{\epsilon_{min}^2}$ becomes

$$\epsilon_{min}^2 = \frac{N_o}{\pi} \int_{-\infty}^{\infty} \left| H_{opt}(\omega) \right| \cos \phi(\omega) \, d\omega.$$

Since $\left| H_{opt}(\omega) \right|$ is an even function and $\phi(\omega)$ is an odd function of ω,

$$\overline{\epsilon_{min}^2} = \frac{N_o}{\pi} \int_{-\infty}^{\infty} \left| H_{opt}(\omega) \right| e^{j\phi(\omega)} \, d\omega$$

$$= \frac{N_o}{\pi} \int_{-\infty}^{\infty} H_{opt}(\omega) \, d\omega \tag{8}$$

$$= N_o \left[h_{opt}(0-) + h_{opt}(0+) \right].$$

Since $h_{opt}(0-) = 0$, Eq. 2 follows. The second expression in Eq. 3 is obtained by use of the initial value theorem. From these expressions, it is also observed that for large s, $H_{opt}(s)$ behaves as $\dfrac{h_{opt}(0+)}{a}$.

3. Derivation of $\overline{\epsilon_{min}^2}$ in Terms of the Signal and Noise Spectra

A simple application of contour integration and the residue theorem show that for any function F(s) that is analytic in the right half-plane (RHP) and behaves as $\dfrac{1}{s^n}$ ($n > 1$) for large s,

$$\int_{-\infty}^{\infty} F(\omega) \, d\omega = 0.$$

Therefore, $\dfrac{N_o}{\pi} \int_{-\infty}^{\infty} \dfrac{1}{n} H_{opt}^n(\omega) \, d\omega = 0$ for $n > 1$ and, from Eq. 8

$$\overline{\epsilon^2_{min}} = \frac{N_o}{\pi} \int_{-\infty}^{\infty} \left[H_{opt}(\omega) + \frac{1}{2} H^2_{opt}(\omega) + \frac{1}{3} H^3_{opt}(\omega) + \dots \right] d\omega.$$

Since $\log(1-z) = -\left(z + \frac{1}{2} z^2 + \frac{1}{3} z^3 + \dots\right)$, it follows that

$$\overline{\epsilon^2_{min}} = -\frac{N_o}{\pi} \int_{-\infty}^{\infty} \log(1-H_{opt}(\omega)) \, d\omega$$

$$= -\frac{N_o}{2\pi} \int_{-\infty}^{\infty} \log(1-H_{opt}(\omega)) \, d\omega - \frac{N_o}{2\pi} \int_{-\infty}^{\infty} \log(1-H_{opt}(-\omega)) \, d\omega \qquad (9)$$

$$= -\frac{N_o}{2\pi} \int_{-\infty}^{\infty} \log\left|1-H_{opt}(\omega)\right|^2 d\omega.$$

The desired relation is obtained upon substitution of Eq. 7 and Eq. 9. The derivation used here is similar to that used by Viterbi and Cahn.[3]

4. Minimum Phase Property of $H_{opt}(s)$

$H_{opt}(s)$ is of the form $1-G(s)$, where $G(s)$ is minimum phase and $\left|G(j\omega)\right| \leq 1$. A Nyquist plot of $H_{opt}(\omega)$ clearly has no encirclements of the origin. Therefore, the number of RHP poles and zeros of $H_{opt}(s)$ is equal. Since $H_{opt}(s)$ has no RHP poles, it has no RHP zeros and is minimum phase. This result is in agreement with the views expressed by Professor H. B. Lee in a discussion which the writer had with him on the problem.

APPENDIX

$$\left|1-H_{opt}(\omega)\right|^2 = (1-H_{opt}(\omega))(1-H_{opt}(-\omega))$$

$$= 1 + \left|H_{opt}(\omega)\right|^2 - 2\left|H_{opt}(\omega)\right| \cos\phi(\omega),$$

where

$$H_{opt}(\omega) = \left|H_{opt}(\omega)\right| e^{j\phi(\omega)}.$$

Using Eq. 7, we have

$$\left|H_{opt}(\omega)\right|^2 = -\frac{N_o S_a(\omega^2)}{S_a(\omega^2) + N_o} + 2\left|H_{opt}(\omega)\right| \cos\phi(\omega). \qquad (10)$$

D. L. Snyder

References

1. M. Yovits and J. Jackson, Linear Filter Optimization with Game Theory Considerations, 1955 IRE National Convention Record, Vol. 3, Part 4. pp. 193-199.

2. J. Develet, Jr., A Threshold Criterion for Phase -Lock Demodulation, Proc. IEEE, Vol. 51, No. 2, pp. 349-356, 1963.

3. A. Viterbi and C. Cahn, Optimum Coherent Phase and Frequency Demodulation of a Class of Modulating Spectra, IEEE Trans. on Space Electronics and Telemetry, Vol. 10, No. 3, pp. 95-102, 1964.

4. H. L. Van Trees and C. Boardman, Optimum Angle Modulation (submitted for publication to IEEE Transactions on Communication Technology).

5. Y. W. Lee, Statistical Theory of Communication (John Wiley and Sons, Inc., New York, 1960).

6. G. Newton, L. Gould, and J. Kaiser, Analytical Design of Feedback Controls (John Wiley and Sons, Inc., New York, 1957).

7. W. Davenport and W. Root, An Introduction to the Theory of Random Signals and Noise (McGraw-Hill Book Company, New York, 1958).

F. AN APPLICATION OF VOLTERRA FUNCTIONAL ANALYSIS TO SHUNT-WOUND COMMUTATOR MACHINES

The present report is a continuation, in the form of two appendices, of the report published in Quarterly Progress Report No. 76 (pages 198-208).

APPENDIX A

Demonstration of Uniqueness

Consider the input-output equation (Eq. 11):

$$\omega(t) = \int_{-\infty}^{\infty} \ldots \int h_f(\tau_1)\, h_a(\tau_2)\, h_\omega(\tau_3)\, v(t-\tau_1-\tau_3)\, v(t-\tau_2-\tau_3)\, d\tau_1 \ldots d\tau_3$$

$$- G \int_{-\infty}^{\infty} \ldots \int h_f(\tau_1)\, h_a(\tau_2)\, h_f(\tau_3)\, h_\omega(\tau_4)\, v(t-\tau_1-\tau_2-\tau_4)\, v(t-\tau_3-\tau_4)$$

$$\cdot\, \omega(t-\tau_2-\tau_4)\, d\tau_1 \ldots d\tau_4$$

For a given input $v(t)$, is the output $\omega(t)$ unique? Let us assume that it is not: Assume that for some $v(t)$ there exist two outputs, $\omega_1(t)$ and $\omega_2(t)$, both of which satisfy Eq. 11. Let the difference between these two outputs be

$$\delta(t) = \omega_1(t) - \omega_2(t). \tag{A.1}$$

Write Eq. 11 for $v(t)$ and $\omega_1(t)$. Write Eq. 11 for $v(t)$ and $\omega_2(t)$. Subtract the latter from the former. The result, with the use of (A.1), is

$$\delta(t) = -G \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} h_f(\tau_1)\, h_a(\tau_2)\, h_f(\tau_3)\, h_\omega(\tau_4)\, v(t-\tau_1-\tau_2-\tau_4)$$

$$\cdot\, v(t-\tau_3-\tau_4)\, \delta(t-\tau_2-\tau_4)\, d\tau_1 \ldots d\tau_4. \tag{A.2}$$

If $v(t)$ exists, then it is everywhere bounded.

$$|v(t)| \leq V, \quad \text{for all } t. \tag{A.3}$$

By assumption, both $\omega_1(t)$ and $\omega_2(t)$ exist and therefore $\delta(t)$ is everywhere bounded.

$$|\delta(t)| \leq D, \quad \text{for all } t. \tag{A.4}$$

Equations A.2, A.3, and A.4 show that

$$|\delta(t)| \leq GV^2 D \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} |h_f(\tau_1)\, h_a(\tau_2)\, h_f(\tau_3)\, h_\omega(\tau_4)|\, d\tau_1 \ldots d\tau_4. \tag{A.5}$$

When the expression for h_f, h_a, h_ω (Eqs. 5, 7, and 10) are substituted in (A. 5), the evaluation of the integrals yields

$$|\delta(t)| \leq \frac{G^2 V^2}{AR_a R_f^2} D.$$

(A. 6)

The following lemma has thus been demonstrated.

LEMMA: If $|v(t)| \leq V$ and $|\delta(t)| \leq D$, then $|\delta(t)| \leq kD$, where

$$k = \frac{G^2 V^2}{AR_a R_f^2}.$$

(A. 7)

This lemma is self-reflexive. That is, since $|\delta(t)| \leq kD$, $|\delta(t)| \leq k^2 D$, and so forth. Thus

$$|\delta(t)| \leq k^m D, \qquad m = 0, 1, 2, \ldots.$$

(A. 8)

If $k < 1$, then as m increases without limit, (A. 8) shows that

$$|\delta(t)| = |\omega_1(t) - \omega_2(t)| = 0.$$

(A. 9)

But, as long as the bound on $v(t)$ satisfies Eq. 12,

$$|v(t)| < \frac{R_f}{G} \sqrt{AR_a}, \quad \text{for all } t,$$

then k is less than one and the solutions to Eq. 11 are unique.

APPENDIX B

Demonstration of Equation 36

Consider Eq. 36:

$$\int_0^\infty \cdots \int |\Omega_{2m}(\tau_1, \ldots, \tau_{2m})| \, d\tau_1 \cdots d\tau_{2m} = \left(\frac{R_f}{G}\right)\left(\frac{G^2}{AR_a R_f^2}\right)^m \quad \text{for } m = 1, 2, 3, \ldots$$

which will be demonstrated by induction.

Initial Case $m=1$

Observe that h_f, h_a, and h_ω (Eqs. 5, 7, and 10) are non-negative. Thus $g(\alpha, \beta)$, which is formed by an integral of their product (Eq. 20), is also non-negative. The kernel $\Omega_2(\tau_1, \tau_2)$ is formed by a summation of g's (Eq. 23) and is, therefore, non-negative. Thus

$$\int_0^\infty \int_0^\infty |\Omega_2(\tau_1, \tau_2)| \, d\tau_1 d\tau_2 = \int_0^\infty \int_0^\infty \Omega_2(\tau_1, \tau_2) \, d\tau_1 d\tau_2. \tag{B.1}$$

But, by Eq. 23,

$$\int_0^\infty \int_0^\infty \Omega_2(\tau_1, \tau_2) \, d\tau_1 d\tau_2 = \int_0^\infty \int_0^\infty g(u, u') \, dudu'. \tag{B.2}$$

By Eq. 20 and suitable change of dummy variables,

$$\int_0^\infty \int_0^\infty g(u, u') \, dudu' = \int_0^\infty \int_0^\infty \int_0^\infty h_\omega(\tau_1) \, h_f(\tau_2) \, h_a(\tau_3) \, d\tau_1 d\tau_2 d\tau_3. \tag{B.3}$$

Finally, substitution of the expressions for h_f, h_a, and h_ω (Eqs. 5, 7, and 10) yields

$$\int_0^\infty \int |\Omega_2(\tau_1, \tau_2)| \, d\tau_1 d\tau_2 = \frac{G}{A} \cdot \frac{1}{R_f} \cdot \frac{1}{R_a}, \tag{B.4}$$

which demonstrates Eq. 36 when m = 1.

Inductive Case m=M+1

Assume that Eq. 36 is true for m = M. We shall then show that it is true for m = M+1: Consider Eq. 24 for the case in which n = 2M + 2. Due to the fact that the SYM operation does not alter the value of this integration, with appropriate change of dummy variables, we have

$$\int_0^\infty \cdots \int |\Omega_{2(M+1)}(\tau_1, \ldots, \tau_{2M+2})| \, d\tau_1 \cdots d\tau_{2M+2} = G \int_0^\infty \cdots \int |h_a(\zeta_1) \, h_\omega(\zeta_2) \, h_f(\zeta_3) \, h_f(\zeta_4)|$$

$$|\Omega_{2M}(\sigma_1, \ldots, \sigma_{2M})| \, d\zeta_1 \cdots d\zeta_4 d\sigma_1 \cdots d\sigma_{2M}$$

$$= G \cdot \frac{1}{R_a} \cdot \frac{G}{A} \cdot \frac{1}{R_f} \cdot \frac{1}{R_f} \cdot \left(\frac{R_f}{G}\right) \cdot \left(\frac{G^2}{AR_a R_f^2}\right)^M$$

$$= \left(\frac{R_f}{G}\right) \left(\frac{G^2}{AR_a R_f^2}\right)^{M+1}. \tag{B.5}$$

Equation B.5 demonstrates that Eq. 36 is true for m = M+1 if it is true for m = M. Since (B.4) demonstrates Eq. 36 for m = 1, it is now clear that Eq. 36 is true for m = 1, 2, 3, ...

R. B. Parente

XIV. PROCESSING AND TRANSMISSION OF INFORMATION*

Prof. F. C. Hennie III
Prof. R. M. Kennedy
Prof. C. E. Shannon
Prof. J. M. Wozencraft
Dr. R. E. Kahn
D. Chase
J. R. Colton

P. M. Ebert
D. D. Falconer
E. F. Ferretti
G. D. Forney, Jr.
C. J. Johnson
J. Max
R. F. McCann

C. W. Niessen
R. E. Olsen
R. Pilc
J. T. Pinkston III
R. N. Spann
W. R. Sutherland
M. G. Taylor

A. LOWER BOUNDS ON THE TAILS OF PROBABILITY DISTRIBUTIONS

Many problems in the Transmission of Information involve the distribution function of the sums of many random variables evaluated far from the mean. In these situations, a direct application of the Central Limit Theorem is virtually useless as an estimate of the distribution function. The so-called Chernov bound,[1] derived here in Eq. 13, turns out to be much more useful both as an upper bound and as an estimate on the far tails of the distribution function. We shall be primarily concerned, however, with deriving lower bounds and asymptotic estimates for the far tails of the distribution function of the sum of independent random variables. A number of the present results, particularly the asymptotic expressions, Eqs. 54 and 61, are due to C. E. Shannon.[2] They are reproduced because of their inaccessibility. The idea of the lower bound in Eq. 74 is also due to Shannon, although the result is stronger here in that it applies to nonidentically distributed variables. Another lower bound to the tail of a distribution has been given by Fano.[3] Fano's approach is to bound the multinomial coefficients for a sum of discrete finite random variables. Our results are more general than Fano's, since they are not restricted to discrete finite variables. On the other hand, in some situations, Fano's bound is tighter than our bounds.

Let ξ be a random variable with the distribution function $F(x) = P(\xi \leq x)$. We shall derive lower bounds to $1 - F(x)$ for $x > \bar{\xi}$, where $\bar{\xi}$ is the expectation of ξ. The bounds will be given in terms of the semi-invariant moment-generating function of ξ,

$$\mu(s) = \ln \int_{-\infty}^{\infty} \exp(sx)\, dF(x) = \ln \overline{\exp s\xi}, \tag{1}$$

in which the bar again denotes expectation.

The bounds will be useful primarily in situations for which ξ is the sum of a sequence of independent random variables, $\xi = \sum_{n=1}^{N} \xi_n$, where each ξ_n has a semi-invariant moment-generating function $\mu_n(s)$. Then

---

$$\mu(s) = \ln \overline{\exp s \sum_{n=1}^{N} \xi_n} = \ln \prod_{n=1}^{N} \overline{\exp s\xi_n} \tag{2}$$

$$= \ln \prod_{n=1}^{N} \overline{\exp s\xi_n} = \sum_{n=1}^{N} \mu_n(s). \tag{3}$$

In going from Eq. 2 to Eq. 3 we have used the fact that for statistically independent random variables, the product of the averages is equal to the average of the product. Equation 3 will allow us to express bounds involving $\mu(s)$ in terms of the $\mu_n(s)$ without explicitly finding the distribution function $F(x)$. The semi-invariant moment-generating function exists for any random variable that takes on only a finite number of values and for any random variable ξ whose probability density drops off _faster_ than exponentially as $\xi \to +\infty$ and as $\xi \to -\infty$. If the probability density drops off only exponentially, then $\mu(s)$ will exist only for a range of s. In the sequel, we assume an $F(x)$ for which $\mu(s)$ exists. If $\mu(s)$ exists only in a region, we consider only values of s in the interior of that region.

In order to find a lower bound to $1 - F(x)$, it is convenient to define a random variable ξ_s with the probability distribution function

$$F_s(x) = \frac{\displaystyle\int_{-\infty}^{x} \exp(sx')\, dF(x')}{\displaystyle\int_{-\infty}^{\infty} \exp(sx')\, dF(x')}. \tag{4}$$

The function $F_s(x)$ is generally called a tilted probability distribution, since it "tilts" the probability assigned by $F(x)$ by the factor e^{sx}. We now show that the mean and variance of the random variable ξ_s for a given s are given by the first and second derivatives of $\mu(s)$ evaluated at the same s. By direct differentiation of Eq. 1, we get

$$\mu'(s) = \frac{\displaystyle\int_{-\infty}^{\infty} x \exp(sx)\, dF(x)}{\displaystyle\int_{-\infty}^{\infty} \exp(sx')\, dF(x')} = \int_{-\infty}^{\infty} x\, dF_s(x) \tag{5}$$

$$\mu''(s) = \frac{\displaystyle\int_{-\infty}^{\infty} x^2 \exp(sx)\, dF(x)}{\displaystyle\int_{-\infty}^{\infty} \exp(sx')\, dF(x')} - [\mu'(s)]^2. \tag{6}$$

Thus

$$\mu'(s) = \overline{\xi_s}; \quad \mu''(s) = \overline{\xi_s^2} - \overline{\xi_s}^2. \tag{7}$$

Since $\mu''(s)$ is a variance and thus strictly positive for nontrivial distributions, $\bar{\xi}_s$ is an increasing function of s.

Using Eq. 4, we now have, for any given s,

$$dF_s(x) = \exp[-\mu(s)+sx]\, dF(x) \tag{8}$$

$$1 - F(x) = \int_{x'>x}^{\infty} dF(x') = \int_{x'>x}^{\infty} \exp[\mu(s)-sx']\, dF_s(x'). \tag{9}$$

It will now be instructive to find an upper bound to $1 - F(x)$ before proceeding to our major objective of lower-bounding $1 - F(x)$. For $s \geqslant 0$, we can upper-bound $\exp(-sx')$ in Eq. 9 by $\exp(-sx)$, and thus obtain

$$1 - F(x) \leqslant \exp[\mu(s)-sx] \int_{x'>x}^{\infty} dF_s(x')$$

$$\leqslant \exp[\mu(s)-sx]\, [F_s(\infty)-F_s(x)]$$

$$\leqslant \exp[\mu(s)-sx]; \quad s \geqslant 0. \tag{10}$$

Since Eq. 10 is valid for any $s \geqslant 0$, we can get the best bound by minimizing $\mu(s) - sx$ with respect to s; if a solution exists for $s \geqslant 0$, it is

$$\mu'(s) = x. \tag{11}$$

Since $\mu''(s) \geqslant 0$, Eq. 11 does indeed minimize $\mu(s) - sx$. Finally, since $\mu'(s)$ is a continuous increasing function of s, we see that a solution will exist for s if

$$\bar{\xi} = \mu'(0) \leqslant x < \lim_{s \to \infty} \mu'(s). \tag{12}$$

Also, it can be seen from Eq. 5 that either $\lim_{s \to \infty} \mu'(s) = \infty$ or $\lim_{s \to \infty} \mu'(s)$ is the smallest x for which $F(x) = 1$, that is, the largest value taken on by the random variable ξ. Substituting Eq. 11 in 10, we get the well-known Chernov bound,[1] given in parametric form,

$$1 - F[\mu'(s)] \leqslant \exp[\mu(s)-s\mu'(s)] \quad s \geqslant 0. \tag{13}$$

The exponent in Eq. 13, $\mu(s) - s\mu'(s)$, is zero for $s = 0$ and has a derivative of $-s\mu''(s)$ with respect to s. Thus for nontrivial distributions, the exponent is negative for $s > 0$. Figure XIV-1 gives a graphical interpretation of the terms in Eq. 13 for a typical random variable.

If we substitute Eq. 3 in Eq. 13, we obtain

$$1 - F\left[\sum_{n=1}^{N} \mu'_n(s)\right] \leqslant \exp \sum_{n=1}^{N} [\mu_n(s)-s\mu'_n(s)]. \tag{14}$$
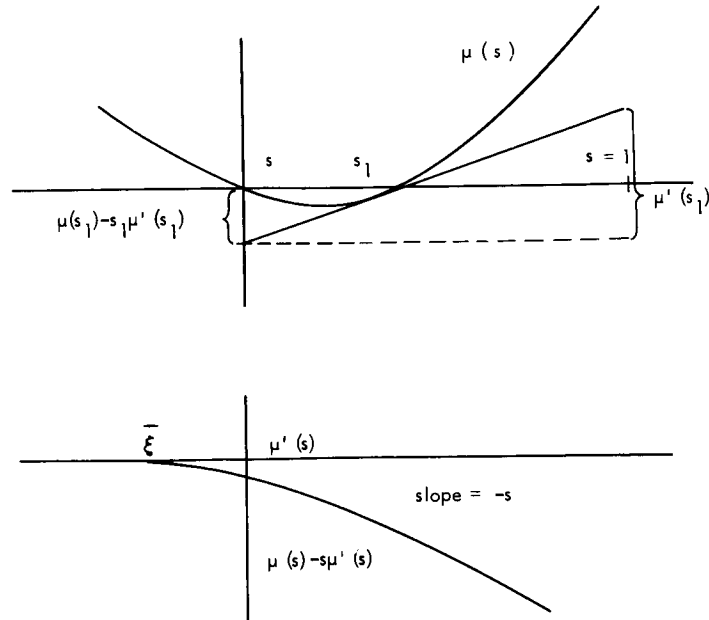
Fig. XIV-1. Graphical interpretations of Eq. 13.

If the ξ_n are all identically distributed, then the argument on the left and the exponent on the right are linear in N.

We next turn our attention to finding a lower bound to $1 - F(x)$. Since x is arbitrary in (9), let us substitute $\mu'(s) - A$ for x, where A is an arbitrary positive number to be chosen later.

$$1 - F[\mu'(s)-A] = \int_{x'>\mu'(s)-A}^{\infty} \exp[\mu(s)-sx'] \, dF_s(x')$$

$$\geq \int_{-A<x-\mu'(s)<A} \exp[\mu(s)-sx] \, dF_s(x). \tag{15}$$

Here, we have lower-bounded the left side by reducing the interval of integration. Restricting s to be non-negative, we observe that $\exp(-sx)$ is decreasing with x and is lower-bounded in (15) by $\exp[-s\mu'(s)-sA]$.

$$1 - F[\mu'(s)-A] \geq \exp[\mu(s)-s\mu'(s)-sA] \int dF_s(x) \qquad -A < x - \mu'(s) < A. \tag{16}$$

Recalling that ξ_s has a mean $\mu'(s)$ and a variance $\mu''(s)$, we can lower-bound the integral in (16) by the Chebyshev inequality,

$$\int dF_s(x) \geq 1 - \frac{\mu''(s)}{A^2} \qquad -A < x - \mu'(s) < A. \tag{17}$$

Choosing $A = \sqrt{2\mu''(s)}$, for simplicity, and substituting (17) in (16), we get

$$1 - F[\mu'(s)-\sqrt{2\mu''(s)}] \geq \frac{1}{2} \exp[\mu(s)-s\mu'(s)-s\sqrt{2\mu''(s)}]. \tag{18}$$

It is convenient to simplify the left-hand side of (18) at the expense of the right-hand side. Define s_1 to satisfy

$$\mu'(s_1) = \mu'(s) - \sqrt{2\mu''(s)}. \tag{19}$$

Expanding $\mu(s)$ in a Taylor expansion around $\mu(s_1)$, we get

$$\mu(s) = \mu(s_1) + (s-s_1)\mu'(s_1) + \frac{(s-s_1)^2}{2}\mu''(r); \quad s_1 \leq r \leq s$$

$$\mu(s) \geq \mu(s_1) + (s-s_1)\mu'(s_1). \tag{20}$$

Substituting (19) and (20) in (18), we have

$$1 - F[\mu'(s_1)] \geq \frac{1}{2} \exp[\mu(s_1)-s_1\mu'(s_1)-2s\sqrt{2\mu''(s)}], \tag{21}$$

where $s \geq 0$, and s_1 is related to s through Eq. 19. Observe that Eqs. 13 and 21 are quite closely related. They differ primarily in the term $2s\sqrt{2\mu''(s)}$. When ξ is the sum of independent random variables, we see from Eq. 3 that $\sqrt{2\mu''(s)}$ is proportional to the square root of the number of random variables, whereas $\mu(s_1)$ and $s_1\mu'(s_1)$ are directly proportional to the number of variables. Thus, in some sense, $2s\sqrt{2\mu''(s)}$ should be unimportant for large N. Unfortunately, giving a precise meaning to this is somewhat involved as the next theorem illustrates.

THEOREM 1: Let ξ_1, ξ_2, ... be an infinite sequence of random variables with semi-invariant moment-generating functions $\mu_1(s)$, $\mu_2(s)$, ... For any positive number A assume that positive numbers L(A) and U(A) exist such that

$$L(A) \leq \frac{1}{N} \sum_{n=1}^{N} \mu_n''(s) \leq U(A); \quad \text{for all } N \geq 1 \text{ and all } s, \ 0 \leq s \leq A. \tag{22}$$

Then for any $s_1 \geq 0$,

$$\lim_{N \to \infty} \frac{1}{N} \ln \{1-F[\mu'(s_1)]\} = \lim_{N \to \infty} \frac{[\mu(s_1)-s_1\mu'(s_1)]}{N}, \tag{23}$$

where $\xi = \sum_{n=1}^{N} \xi_n$, and F(x) and $\mu(s)$ are the distribution function and semi-invariant moment-generating function of ξ.

DISCUSSION: The condition in Eq. 22 is broad enough to cover situations in which

each of the ξ_n has one of a finite set of distribution functions, each with nonzero variance and a semi-invariant moment-generating function. The number A is brought in to avoid ruling out the broad class of random variables for which $F_n(x) = 1$ for some finite x and thus $\lim_{s\to\infty} \mu_n''(s) = 0$.

PROOF: It follows immediately from (14) that the left side of (23) is less than or equal to the right side. Also, from (21), for any given N, we have

$$\frac{1}{N} \ln \left\{1 - F[\mu'(s_1)]\right\} \geq \frac{\mu(s_1) - s_1\mu'(s_1)}{N} - \frac{\ln 2}{N} - \frac{2s\sqrt{2\mu''(s)}}{N}. \tag{24}$$

In Eq. 24 s and s_1 are related by Eq. 19. Also, from the mean value theorem,

$$\mu'(s_1) = \mu'(s) + (s_1 - s)\mu''(s_2); \quad \text{for some } s_2, \ s_1 \leq s_2 \leq s. \tag{25}$$

Combining Eqs. 19 and 25, we have

$$s - s_1 = \frac{\sqrt{2\mu''(s)}}{\mu''(s_2)}. \tag{26}$$

For any given A > 0 and all s ≤ A, we can upper-bound $\mu''(s)$ and lower-bound $\mu''(s_2)$ by Eq. 22.

$$s - s_1 \leq \frac{\sqrt{2NU(A)}}{NL(A)} = \frac{1}{\sqrt{N}} \frac{\sqrt{2U(A)}}{L(A)}. \tag{27}$$

Next, let ϵ be an arbitrary positive number and restrict N to satisfy

$$N \geq \frac{2U(A)}{[L(A)]^2 \epsilon^2}. \tag{28}$$

From Eqs. 27 and 28,

$$0 \leq s - s_1 \leq \epsilon. \tag{29}$$

Since s_1 is a continuous function of s, s_1 will take on all values from 0 to A − ϵ as s goes from 0 to A. Thus for any s_1, $0 \leq s_1 \leq A - \epsilon$, and for any N satisfying (28), we have s ≤ A, and therefore from Eq. 22

$$\frac{2s\sqrt{2\mu''(s)}}{N} \leq 2s\sqrt{\frac{2U(A)}{N}} \xrightarrow[N\to\infty]{} 0. \tag{30}$$

But, since A and ϵ are arbitrary, Eq. 30 is valid for any s_1. Thus, from (24),

$$\lim_{N\to\infty} \frac{1}{N} \ln \left\{1 - F[\mu'(s_1)]\right\} \geq \lim_{N\to\infty} \frac{\mu(s_1) - s_1\mu'(s_1)}{N}, \tag{31}$$

thereby completing the proof.

It is frequently convenient to have a specific upper bound on $2s\sqrt{2\mu''(s)}$, since it is the essential difference between our upper and lower bounds. The following result is quite crude and applies only to discrete random variables taking on a maximum value. Let the random variable ξ_n take on the values $x_{n1} \geq x_{n2} \geq x_{n3} \geq \ldots$ with probabilities $p_{n1}, p_{n2} \ldots$. From Eq. 6 we know that $\mu_n''(s)$ is the variance of a random variable taking on the values $x_{n1}, x_{n2}, \ldots$ with probabilities $p_{n1} e^{sx_{n1}} \Big/ \sum_i p_{ni} e^{sx_{ni}}, \ldots$. Since the variance of a random variable is upper-bounded by the second moment around any value, we have

$$\mu_n''(s) \leq \sum_k (x_{nk}-x_{n1})^2 \frac{p_{nk} e^{sx_{nk}}}{\sum_i p_{ni} e^{sx_{ni}}}. \tag{32}$$

Multiplying numerator and denominator by $e^{-sx_{n1}}$ and defining η_{nk} by $s(x_{nk}-x_{n1})$, we have

$$s^2 \mu_n''(s) \leq \sum_k \eta_{nk}^2 \frac{p_{nk} e^{\eta_{nk}}}{\sum_i p_{ni} e^{\eta_{ni}}}. \tag{33}$$

Now, $\eta_{n1} = 0$ and $\eta_{nk} \leq 0$ for $s \geq 0$. Thus the denominator in (33) can be lower-bounded by p_{n1}. Furthermore, $\eta_{nk}^2 e^{\eta_{nk}} \leq (2/e)^2$ for any $\eta_{nk} \leq 0$. Incorporating these results in (33) yields

$$s^2 \mu_n''(s) \leq (2/e)^2 \left[\frac{1 - p_{n1}}{p_{n1}} \right], \tag{34}$$

where p_{n1} is the probability of the largest value taken on by ξ_n. Let $p_1 = \min_n p_{n1}$, and use Eq. 3, then, we get

$$s^2 \mu''(s) \leq N(2/e)^2 \left[\frac{1 - p_1}{p_1} \right]. \tag{35}$$

Substituting (35) in (21), we get

$$1 - F[\mu'(s_1)] \geq \frac{1}{2} \exp\left[\mu(s_1) - s_1\mu'(s_1) - \frac{4}{e} \sqrt{\frac{2N(1-p_1)}{p_1}} \right]. \tag{36}$$

The previous results have all been derived through the use of the Chebyshev inequality and are characteristically simple and general but weak. We now turn to the use of

the Central Limit Theorem to get tighter results. As before, the random variable ξ is the sum of N independent random variables, $\xi = \sum_{n=1}^{N} \xi_n$. The ξ_n have the distribution functions $F_n(x)$ and the semi-invariant moment-generating functions $\mu_n(s)$. Now, let us define the tilted random variables $\xi_{n,s}$ with the distribution functions

$$F_{n,s}(x) = \frac{\int_{-\infty}^{x} \exp(sx')\, dF_n(x')}{\int_{-\infty}^{\infty} \exp(sx')\, dF_n(x')}. \tag{37}$$

The semi-invariant moment-generating function of $\xi_{n,s}$ is

$$h_{n,s}(r) = \int_{-\infty}^{\infty} \exp(rx)\, dF_{n,s}(x) \tag{38}$$

$$= \mu_n(s+r) - \mu_n(s), \tag{39}$$

where Eq. 39 follows from substituting (37) in (38).

If ξ_s is now defined as $\xi_s = \sum_{n=1}^{N} \xi_{n,s}$, the semi-invariant moment-generating function of ξ_s is

$$h_s(r) = \sum_{n=1}^{N} h_{n,s}(r) = \mu(s+r) - \mu(s). \tag{40}$$

Thus, if we work backwards, the distribution function of ξ_s is given by Eq. 4.

Now, let $x = \mu'(s)$ in Eq. 9,

$$1 - F[\mu'(s)] = \exp[\mu(s)-s\mu'(s)] \int_{x>\mu'(s)} \exp\{s[\mu'(s)-x]\}\, dF_s(x). \tag{41}$$

We shall assume, temporarily, that the ξ_n are all identically distributed and non-lattice. (A lattice distribution is a distribution in which the allowable values of ξ_n can be written in the form $x_k = hk + a$, where h and a are arbitrary numbers independent of the integer k. The largest h for which the allowable values of ξ_n can be expressed in this way is called the span of the distribution.) Then ξ_s has the mean $\mu'(s) = N\mu_n'(s)$ and the variance $\mu''(s) = N\mu_n''(s)$, and for $s > 0$, the terms in the integral of (41) have the appearance shown in Fig. XIV-2. Observe that $F_s(x)$ is approximately a Gaussian distribution function, but the exponential term is changing much more rapidly than $F_s(x)$ for large N. Let η_s be the normalized random variable.

$$\eta_s = \frac{\xi_s - \mu'(s)}{\sqrt{\mu''(s)}}. \tag{42}$$

Let $G_s(y)$ be the distribution function of η_s,

$$G_s(y) = F_s[y\sqrt{\mu''(s)}+\mu(s)].$$ (43)

Transforming the variable of integration in (41) by (42), we have

$$1 - F[\mu'(s)] = \exp[\mu(s)-s\mu'(s)] \int_{z>0} \exp[-s\sqrt{\mu''(s)}\, z]\, dG(z).$$ (44)

Assuming $s > 0$, we can use integration by parts on the integral in (44), to obtain

$$\int_{z>0} \exp[-s\sqrt{\mu''(s)}\, z]\, dG(z) = s\sqrt{\mu''(s)} \int_{z=0}^{\infty} [G(z)-G(0)]\exp[-s\sqrt{\mu''(s)}\, z]\, dz.$$ (45)

Equation 45 is now in a form suitable for application to the central limit theorem. Since the exponential term is decaying so rapidly, we need a particularly strong central
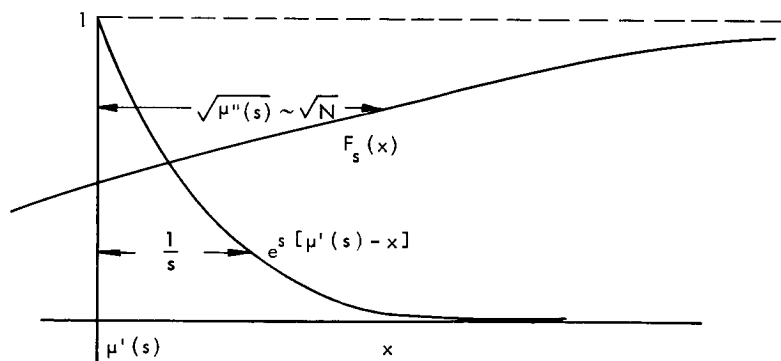
Fig. XIV-2. Sketch of terms in Eq. 41.

limit theorem. The appropriate theorem is due to Esseen[4] and is given by Gnedenko and Kolmogoroff.[5] Under conditions less restrictive than those that we have already assumed, the theorem states

$$G(z) = \Phi(z) + \frac{Q_1(z)\exp\left(-\dfrac{z^2}{2}\right)}{\sqrt{2\pi N}} + o\left(\frac{1}{\sqrt{N}}\right)$$ (46)

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)dx$$ (47)

$$Q_1(z) = Q_1(0) \ (1-z^2) \tag{48}$$

$$Q_1(0) = \frac{\mu_n'''(s)}{6[\mu_n''(s)]^{3/2}} \tag{49}$$

and $o\left(\dfrac{1}{\sqrt{N}}\right)$ is a function approaching 0, uniformly in z, faster than $\dfrac{1}{\sqrt{N}}$; that is, $f(N) = o\left(\dfrac{1}{\sqrt{N}}\right)$ if $\lim\limits_{N\to\infty} \sqrt{N} \ f(N) = 0.$

Substituting (46) and (48) in (45), we have

$$\int\limits_{z>0} \exp[-s\sqrt{\mu''(s)} \ z] \ dG(z)$$

$$= s\sqrt{\mu''(s)} \int_{z=0}^{\infty} [\Phi(z)-\Phi(0)] \exp[-s\sqrt{\mu''(s)} \ z] \ dz$$

$$+ \frac{s\sqrt{\mu''(s)} \ Q_1(0)}{\sqrt{N}} \int_0^{\infty} \left[1 - \exp\left(-\frac{z^2}{2}\right)\right] \exp[-s\sqrt{\mu''(s)} \ z] \ dz$$

$$+ \frac{s\sqrt{\mu''(s)} \ Q_1(0)}{\sqrt{N}} \int_0^{\infty} z^2 \exp\left(-\frac{z^2}{2}\right) \exp[-s\sqrt{\mu''(s)} \ z] \ dz$$

$$+ s\sqrt{\mu''(s)} \int_0^{\infty} o\left(\frac{1}{\sqrt{N}}\right) \exp[-s\sqrt{\mu''(s)} \ z] \ dz. \tag{50}$$

The first integral on the right-hand side of (50) can be integrated by parts, and then by completing the square in the exponent. This yields

$$s\sqrt{\mu''(s)} \int_0^{\infty} [\Phi(z)-\Phi(0)] \exp[-s\sqrt{\mu''(s)} \ z] \ dz = \exp\left[\frac{s^2\mu''(s)}{2}\right] [1-\Phi(s\sqrt{\mu''(s)})] \tag{51}$$

Using standard inequalities on the normal distribution function (see Feller[6]), we obtain

$$\left[1 - \frac{1}{s^2\mu''(s)}\right] \frac{\exp\left[-\dfrac{s^2\mu''(s)}{2}\right]}{\sqrt{2\pi s^2\mu''(s)}} \leq 1 - \Phi(s\sqrt{\mu''(s)}) \leq \frac{\exp\left[-\dfrac{s^2\mu''(s)}{2}\right]}{\sqrt{2\pi s^2\mu''(s)}}. \tag{52}$$

Recalling that $\mu''(s) = N\mu_n''(s)$, we see that the first integral in (50) is equal to $\dfrac{1}{\sqrt{2\pi N s^2\mu_n''(s)}} + o\left(\dfrac{1}{\sqrt{N}}\right).$

A similar integration on the second integral in (50) shows that it can be represented

by $o(1/\sqrt{N})$. If we upper-bound $\exp[-z^2/2]$ in the third integral of (50) by 1, it follows that it also is $o(1/\sqrt{N})$. By using the uniform convergence in z of the $o(1/\sqrt{N})$ in the fourth integral, we see that it too is $o(1/\sqrt{N})$. Thus,

$$\int_{z>0}^{\infty} \exp[-s\sqrt{\mu''(s)}\, z]\, dG(z) = \frac{1}{\sqrt{2\pi N s^2 \mu_n''(s)}} + o\left(\frac{1}{\sqrt{N}}\right). \tag{53}$$

Substituting (53) in (44), we see that for identically distributed nonlattice variables with $s > 0$,

$$1 - F[\mu'(s)] = \left[\frac{1}{\sqrt{2\pi N s^2 \mu_n''(s)}} + o\left(\frac{1}{\sqrt{N}}\right)\right] \exp[\mu(s) - s\mu'(s)]. \tag{54}$$

We shall now derive a relationship similar to Eq. 54 for the lattice case. Let the ξ_n be independent and identically distributed and take on only the values

$$x_k = a + hk, \tag{55}$$

where a and h are arbitrary numbers independent of the integer k, and one is the greatest common divisor of the integers k for which x_k has nonzero probability. The random variables ξ and ξ_s also can take on only the values $Na + hk$ for integer k. Let Δ be the magnitude of the difference between $\mu'(s)$ and the smallest value of $Na + hk$ larger than $\mu'(s)$. Define

$$p_s(j) = P(\xi_s = \mu'(s) + \Delta + hj). \tag{56}$$

We can now apply a central limit theorem for lattice distributions (see Gnedenko and Kolmogoroff[7]). This theorem states, in effect, that for any $\epsilon > 0$, there exists an N_o such that for $N \geqslant N_o$, we have

$$\left| p_s(j) - \frac{h}{\sqrt{2\pi\mu''(s)}} \exp\left[\frac{-(\Delta+jh)^2}{2\mu''(s)}\right] \right| \leqslant \frac{\epsilon}{\sqrt{\mu''(s)}}. \tag{57}$$

Bounding the exponential term, we can rewrite this as

$$-\frac{\epsilon}{\sqrt{\mu''(s)}} + \frac{h}{\sqrt{2\pi\mu''(s)}}\left[1 - \frac{(\Delta+jh)^2}{2\mu''(s)}\right] \leqslant p_s(j) \leqslant \frac{\epsilon}{\sqrt{\mu''(s)}} + \frac{h}{\sqrt{2\pi\mu''(s)}}. \tag{58}$$

Equation 41 can now be rewritten in terms of the $p_s(j)$ to yield

$$1 - F[\mu'(s)] = \exp[\mu(s) - s\mu'(s)] \sum_{j=0}^{\infty} p_s(j) \exp[-s(\Delta+jh)] \tag{59}$$

$$1 - F[\mu'(s)] \leq \exp[\mu(s)-s\mu'(s)]\left[\frac{\epsilon}{\sqrt{\mu''(s)}} + \frac{h}{\sqrt{2\pi\mu''(s)}}\right]\frac{\exp(-s\Delta)}{1 - \exp(-sh)}. \tag{60}$$

Here, we have upper-bounded (59) by (58) and summed over j.

Equation 58 can also be used to lower-bound (59). If we define

$$A = \sum_{j=0}^{\infty} (\Delta+jh)^2 \exp[-s(\Delta+jh)],$$

then

$$1 - F[\mu'(s)] \geq \frac{\exp[\mu(s)-s\mu'(s)]}{\sqrt{2\pi\mu''(s)}}\left[(h-\epsilon\sqrt{2\pi})\frac{\exp(-s\Delta)}{1 - \exp(-sh)} - \frac{hA}{\mu''(s)}\right].$$

Observing that ϵ can be made to approach 0 with increasing N and that A is bounded independently of N, we have

$$1 - F[\mu'(s)] = \left[\frac{h\exp(-s\Delta)}{2\pi N\mu_n''(s)[1-\exp(-sh)]} + o\left(\frac{1}{\sqrt{N}}\right)\right]\exp[\mu(s)-s\mu'(s)]. \tag{61}$$

Equation 61 is valid for any s > 0 for independent, identically distributed lattice variables if $\mu_n(s)$ exists. Note, however, that Δ will fluctuate between 0 and h as a function of N.

Equations 54 and 61 are not applicable in general to nonidentically distributed random variables. In some cases, however, Eqs. 54 and 61 can be made to apply, first, by grouping the variables to make them identically distributed. For example, for N variables, if N/2 variables have one distribution and N/2 have another distribution, then we can form N/2 identically distributed variables, each of which is the sum of a pair of the original variables.

In the sequel, we shall take a different approach to nonidentically distributed variables and derive a lower bound for $1 - F[\mu'(s)]$ by using a different form of the central limit theorem. This new result will be more complicated than (54) and (61), but will have the advantage of providing a firm lower bound to $1 - F[\mu'(s)]$ and of being applicable to nonidentically distributed variables. It will only be stronger than Eq. 21 for large N. We start with Eqs. 44 and 45, which are still valid for nonidentically distributed independent variables. Then the Berry theorem[8] states

$$\left|G(z)-\Phi(z)\right| \leq \frac{C\rho_{3,N}}{\sqrt{N}}, \tag{62}$$

where

$$\rho_{3,N} = \frac{\frac{1}{N} \sum_{n=1}^{N} \beta_{3,n}}{\left[\frac{1}{N} \sum_{n=1}^{N} \mu_n''(s)\right]^{3/2}} \tag{63}$$

$$\beta_{3,n} = \int_{-\infty}^{\infty} |x-\mu_n'(s)|^3 \, dF_{s,n}(x) \tag{64}$$

and C is a constant. Esseen[4] has shown that C may be taken to be 7.5, but no example has ever been found in which C need be larger than 0.41. The constant must be at least 0.41 to cope with a sum of binary random variables, each of which takes on one value with the probability $(\sqrt{10}-2)/2$.

From Eq. 62, and from the fact that G(z) is a distribution function, we have

$$G(z) - G(0) \geq \Phi(z) - \Phi(0) - \frac{2C\rho_{3,N}}{\sqrt{N}} \tag{65}$$

$$\geq 0; \quad z \geq 0. \tag{66}$$

Let z_0 be that value of z for which the right-hand side of (65) is 0.

$$\Phi(z_0) = \frac{1}{2} + \frac{2C\rho_{3,N}}{\sqrt{N}}. \tag{67}$$

Observe that if N is not sufficiently large, (67) will have no solution. More precisely, N must be greater than $[4C\rho_{3,N}]^2$ for Eq. 67 to have a solution. For smaller values of N, we must use Eq. 21 to lower-bound $1 - F[\mu'(s)]$. Because of the importance of $\rho_{3,N}$ here, it is sometimes convenient to have a bound on $\rho_{3,N}$ in terms of $\mu(s)$. Using the theorem of the means, we have

$$\beta_{3,N} \leq \left[\int_{-\infty}^{\infty} [x-\mu_n'(s)]^4 \, dF_{n,s}(x)\right]^{3/4} = \{\mu_n''''(s)+3[\mu_n''(s)]^2\}^{3/4} \tag{68}$$

$$\rho_{3,N} \leq \frac{\frac{1}{N} \sum_{n=1}^{N} \{\mu_n''''(s)+3[\mu_n''(s)]^2\}^{3/4}}{\left[\frac{1}{N} \sum_{n=1}^{N} \mu_n''(s)\right]^{3/2}}. \tag{69}$$

Using Eq. 66 for $z < z_0$, we find that Eq. 45 becomes

$$\int_{z>0} \exp[-s\sqrt{\mu''(s)}\, z] \, dG(z) \geq s\sqrt{\mu''(s)} \int_{z_0}^{\infty} \left[\Phi(z) - \Phi(0) - \frac{2C\rho_{3,N}}{\sqrt{N}}\right] \exp[-s\sqrt{\mu''(s)}\, z] \, dz. \tag{70}$$

Integrating by parts and using Eq. 67, we get

$$\int_{z>0} \exp[-s\sqrt{\mu''(s)}\,z]\,dG(z) \geq \int_{z_0}^{\infty} \exp[-s\sqrt{\mu''(s)}\,z]\,d\Phi(z) \tag{71}$$

$$\geq \exp\left[\frac{s^2\mu''(s)}{2}\right]\left[1 - \Phi(z_0 + s\sqrt{\mu''(s)})\right] \tag{72}$$

$$\geq \frac{\exp\left[-z_0 s\sqrt{\mu''(s)} - \frac{z_0^2}{2}\right]}{\sqrt{2\pi}\,[z_0 + s\sqrt{\mu''(s)}\,]}\left\{1 - \frac{1}{[z_0 + s\sqrt{\mu''(s)}\,]^2}\right\}. \tag{73}$$

Equation 71 was integrated by completing the square in the exponent, and in Eq. 73 we used the bound on $\Phi(x)$ given by Eq. 52. If we define B as the right-hand side of Eq. 73, then from Eq. 44, we have

$$1 - F[\mu'(s)] \geq B \exp[\mu(s) - s\mu'(s)]. \tag{74}$$

It is instructive to estimate B for very large N, under the assumption that $\mu''(s)$ grows in some sense linearly with N. Under these circumstances, from Eq. 67,

$$z_0 \approx \frac{2C\rho_{3,N}}{\sqrt{2\pi N}} \tag{75}$$

$$B \approx \frac{\exp\left[-2Cs\rho_{3,N}\sqrt{\sum_{n=1}^{N}\frac{\mu_n''(s)}{N}}\right]}{\sqrt{2\pi s^2 \sum_{n=1}^{N}\mu_n''(s)}}. \tag{76}$$

We see that for large N, Eqs. 54 and 74 differ by the numerator of Eq. 76. This term is essentially independent of N, but is typically very small relative to 1.

All of the results thus far are concerned with the upper tail of a distribution function, $1 - F(x)$, for $x > \bar{\xi}$. We can apply all of these results to the lower tail of a distribution, $F(x)$, for $x < \bar{\xi}$, simply by considering the random variable $-\xi$ rather than ξ. Since the semi-invariant moment-generating function of $-\xi$ is related to that of ξ through a change in sign of s, we can write the results immediately in terms of $\mu(s)$ for $s < 0$. Equation 13 becomes

$$F[\mu'(s)] \leq \exp[\mu(s) - s\mu'(s)]. \tag{77}$$

(Actually $F[\mu'(s)] = P[\xi \leq \mu'(s)]$, whereas the counterpart of Eq. 13 treats $P[\xi < \mu'(s)]$. A trivial modification of Eqs. 9-13 establishes the stronger result.)

Upon recognizing that $F[\mu'(s)] \geq P[\xi < \mu'(s)]$, Eqs. 18, 19, and 21 become

$$F[\mu'(s)+2\sqrt{\mu''(s)}\,] \geq \tfrac{1}{2}\,\exp[\mu(s)-s\mu'(s)+s\sqrt{2\mu''(s)}\,] \tag{78}$$

$$\mu'(s_1) = \mu'(s) + \sqrt{2\mu''(s)} \tag{79}$$

$$F[\mu'(s_1)] \geq \tfrac{1}{2}\,\exp[\mu(s_1)-s_1\mu'(s_1)+2s\sqrt{2\mu''(s)}\,]. \tag{80}$$

Equation 54, for identically distributed nonlattice variables, is

$$F[\mu'(s)] = \left[\frac{1}{\sqrt{2\pi N s^2 \mu_n''(s)}} + o\!\left(\frac{1}{\sqrt{N}}\right)\right]\exp[\mu(s)-s\mu'(s)]. \tag{81}$$

Equation 61, for identically distributed lattice variables, is

$$F[\mu'(s)] = \left\{\frac{h\,\exp(\ s\Delta)}{\sqrt{2\pi N\mu_n''(s)}\,[1-\exp(\ sh)]} + o\!\left(\frac{1}{\sqrt{N}}\right)\right\}\exp[\mu(s)-s\mu'(s)], \tag{82}$$

where Δ is the interval between $\mu'(s)$, and the largest value of $Na + kh$ less than or equal to $\mu'(s)$. Finally, Eq. 74 becomes

$$F[\mu'(s)] \geq B\,\exp[\mu(s)-s\mu'(s)] \tag{83}$$

$$B = \frac{\exp\!\left[z_0 s\sqrt{\mu''(s)} - \dfrac{z_0^2}{2}\right]}{\sqrt{2\pi}\,[z_0-s\sqrt{\mu''(s)}\,]}\left\{1 - \frac{1}{[z_0-s\sqrt{\mu''(s)}\,]^2}\right\}, \tag{84}$$

where z_0 is given by Eq. 67, and Eqs. 77-84 are valid for $s < 0$.

R. G. Gallager

References

1. H. Chernov, A measure of asymptotic efficiency for tests of an hypothesis based on the sum of observations, Ann. Math. Statist. 23, 493 (1952).

2. C. E. Shannon, Seminar Notes for Seminar in Information Theory at M. I. T., 1956 (unpublished).

3. R. M. Fano, Transmission of Information (The M. I. T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961); see Chap. 8.

4. C. G. Esseen, Fourier analysis of distribution functions: A mathematical study of the Laplace-Gaussian law, Acta Math. 77, 1-125 (1945).

5. B. V. Gnedenko and A. N. Kolmogoroff, Limit Distributions for Sums of Independent Random Variables (Addison-Wesley Publishing Company, Inc., Cambridge, Mass., 1954); see p. 210.

6. W. Feller, Probability Theory and Its Applications (John Wiley and Sons, Inc., New York, 1950); see Chap. 7.

7. B. V. Gnedenko and A. N. Kolmogoroff, op. cit., p. 233.

8. A. C. Berry, The accuracy of the Gaussian approximation to the sum of independent variates, Trans. Am. Math. Soc. 49, 122-136 (1941).

B. ERROR BOUNDS FOR GAUSSIAN NOISE CHANNELS

Considerable work on estimating the achievable probability of error in communication over a wide variety of channels has indicated that for time-discrete channels the minimum achievable error probability has the form

$$P_e \approx e^{-NE(R)},$$

where N is the block length on which coding and decoding operations are carried out, R is the transmission rate, in nats per channel use, and E(R) is a function of R and of the channel but is independent of N. The approximation is usually close only for large values of N. Thus the estimation of the function E(R) amounts to estimating

$$\lim_{N \to \infty} -\frac{\ln P_e}{N}.$$

Usually it is hard to calculate E(R) exactly, and bounds for the function are found instead. The upper and lower bounds are defined as follows: For any $\epsilon > 0$ and sufficiently large N there exists a code for which

$$P_e \leq e^{-N[E_U(R)-\epsilon]}$$

and there exists no code for which:

$$\overline{P}_e \leq e^{-N[E_L(R)+\epsilon]}.$$

Note that $E_U(R)$ is a lower bound to E(R), but it arises in upper-bounding the achievable P_e, and $E_L(R)$ is an upper bound to E(R) used in lower-bounding P_e.

Gallager[1,2] has found a number of these upper and lower bounds for the discrete memoryless channel and also for the time-discrete Gaussian noise channel with a power constraint. Shannon[3] found some of the same bounds for the bandlimited white Gaussian noise channel, except that in his case the block length is replaced by the time duration over which coding and decoding operations take place

$$P_e \leq e^{-TE(R)},$$

and the rate R is in nats per second. If one takes the limit of these bounds for $N \to \infty$, one finds that all of the bounds have some properties in common. The limits of the upper and lower bounds coincide for R = 0 and $R \geq R_{crit}$, where R_{crit} is a function of the channel and lies between 0 and capacity. The bounds are decreasing convex downward functions of R, and all become zero at R = Capacity.

The channel model that is analyzed here consists of a number of time discrete

channels each disturbed by an independent Gaussian noise. Each channel is to be used only once, and the total energy used in signaling over the channels is constrained. There is no limitation on the number of channels, nor is the value of the noise power required to be the same in all channels. This model, therefore, represents one channel used N times, a set of Q parallel channels with arbitrary noise power in each used N times, or parallel channels, all having different noise powers each of which is used once. This model also represents the colored Gaussian noise channel with an average power constraint. One takes a Karhunen-Loeve expansion of the noise over a T-second interval. Each of the eigenfunctions of the noise autocorrelation function so obtained is considered as one of the component channels in the model. When the noise is Gaussian, the Karhunen-Loeve theorem states that the noise power in each of the eigenfunctions is independent, which is exactly what is needed for the model. As T is made large, the distribution of the noises in the eigenfunction channels approaches the power density spectrum of the noise, and the resulting E(R) function can be expressed in terms of this spectrum. In this case, the energy constraint is PT, where P is just the power available.

Techniques similar to those used by Gallager in obtaining upper and lower bounds for the discrete memoryless channel can be applied here, except that now there is some added freedom; the energies distributed to each of the component channels of the model are subject only to the constraint that they be positive and add up to NP on the average. With this freedom comes the new problem of determining the optimum distribution of energy to the component channels.

When the various bounds are evaluated a remarkable phenomenon appears. As might be expected, only the component channels with noise power below a threshold (N_b) are to be used for communication, but the value of the threshold over most of the parameter range is dependent only on the rate, and is independent of the power available or of the probability of error desired.

1. Lower Bound on E(R)

Since our model consists of a number of parallel channels, each of which is used only once, it has an implicit block length of one, and the resulting bound is of the form

$$P_e \leq e^{-E^*(R^*)+\epsilon},$$

where now R^* is the nats per block. The quantity ϵ will be discussed in more detail. Suffice it to say now that when we have Q channels, each used N times, $\dfrac{\epsilon}{E^*(R^*)}$ will go to zero with increasing N. Also, when the channels come from the eigenfunctions of a Karhune-Loeve expansion, $\dfrac{\epsilon}{E^*(R^*)} \rightarrow 0$ as $T \rightarrow \infty$.
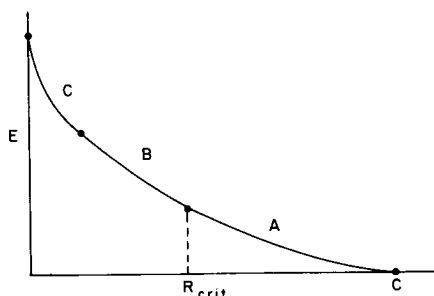
Fig. XIV-3. $E_U(R)$ for fixed power. (Because of the similar relations among R^* and R, E^* and E, and S and P, the curves are drawn for E, R, and P.)

The lower bound takes on three different forms as shown in Fig. XIV-3. Let μ_n be the noise power in the nth channel, and let S be the total energy constraint on the inputs. Then in region A, the relations are

$$R^* = \frac{1}{2} \sum_{N_n \leqslant N_b} \ln \frac{N_b}{N_n}, \qquad \rho \leqslant 1 \tag{1}$$

$$E^* = \frac{\rho S}{(1+\rho)\, 2N_b} - \frac{1}{2} \sum_{N_n \leqslant N_b} \ln \left(1 + \rho - \rho \frac{N_n}{N_b}\right), \tag{2}$$

where

$$S = (1+\rho) \sum_{N_n \leqslant N_b} \frac{N_b - N_n}{1 - \frac{\rho}{1+\rho} \frac{N_n}{N_b}}. \tag{3}$$

For a given rate R^*, and a given energy S, we observe that N_b is defined by Eq. 1, ρ by Eq. 3, and E^* by Eq. 2; the bound is valid in the region where the resultant ρ lies in the interval $(0, 1)$.

The form of Eqs. 1, 2, and 3 is somewhat different from Gallager and Shannon's, in that it has two parameters (ρ, N_b) rather than the usual one. This is not a serious problem if one approaches it in a slightly different manner. Instead of specifying the energy and then finding R^* and E^* as functions of ρ, we first specify the rate R^*. R^* determines N_b; although this is not a simple relation, it is a one-to-one relation. Once N_b is determined, one has E^* and S as functions of the parameter ρ.

To see how the probability of error goes to zero with increasing N for the Q-channel case, we note that the number of component channels in the model with a given value of N_n is a multiple of N, and consequently

$$\frac{1}{2} \sum_{N_n \leq N_b} \frac{N_b - N_n}{1 - \dfrac{\rho}{1 + \rho} \dfrac{N_n}{N_b}} = \frac{N}{2} \sum_{N_q \leq N_b} \frac{N_b - N_q}{1 - \dfrac{\rho}{1 + \rho} \dfrac{N_q}{N_b}}$$

in which the sum on the right is just over those of the Q original channels that have noise power equal to or less than N_b. The same thing is true of the other sums, and we can write

$$R^* = NR = \frac{N}{2} \sum_{N_q \leq N_b} \ln \frac{N_b}{N_q}$$

$$E^* = NE = \frac{N\rho P}{2(1+\rho) N_b} - \frac{N}{2} \sum_{N_q \leq N_b} \ln \left(1 + \rho - \rho \frac{N_q}{N_b} \right), \tag{4}$$

where

$$S = NP = (1+\rho)N \sum_{N_n \leq N_b} \frac{N_b - N_q}{1 - \dfrac{\rho}{1 + \rho} \dfrac{N_q}{N_b}}.$$

For the colored-noise case, as $T \to \infty$ it can be proved that

$$\frac{1}{2} \sum_{N_n \leq N_b} \ln \frac{N_b}{N_n} \to T \cdot \frac{1}{2} \int_{N(f) \leq N_b} \ln \frac{N(f)}{N_b} df + 0(T),$$

where $\dfrac{0(T)}{T} \to 0$, and $N(f)$ is the power density spectrum of the noise. The same thing is true of the other summations, and relations similar to (4) can be written. The boundary of region A is set by $\rho = 1$.

In region C we have

$$R^* = \frac{1}{2} \sum_{N_n \leq N_b} \ln \frac{N_b}{N_n} - \frac{1}{2} \sum_{N_n \leq N_b} \ln \left(2 - \frac{N_n}{N_b} \right) \tag{5}$$

$$E^* = \frac{S}{4N_b}.$$

For completeness we write

$$S = 4\rho \sum_{N_n \leq N_b} \frac{N_b - N_n}{2 - \dfrac{N_n}{N_b}} \qquad \rho \geq 1.$$

In both regions $\dfrac{\partial E^*}{\partial R^*} = \dfrac{\partial E}{\partial R} = -\rho.$

The functions in region B are found to be

$$S = 4 \sum_{N_n \leqslant N_b} \frac{N_b - N_n}{2 - \dfrac{N_n}{N_b}}$$

$$E^* = \frac{S}{4N_b} + \frac{1}{2} \sum_{N_n \leqslant N_b} \ln \frac{N_b}{N_n} - \frac{1}{2} \sum_{N_n \leqslant N_b} \ln \left(2 - \frac{N_n}{N_b}\right) - R^*.$$

In this region ρ is held at 1, and consequently $\dfrac{\partial E^*}{\partial R} = -1$; thus only the variable N_b is left to adjust the trade-off between E^* and S. In this region N_b is not a function of R^*, but of S. The separations of regions A, B, and C can be made by examining the value of ρ. Once the values of N_b and ρ have been determined, it is a simple matter to design the signals that will produce a probability of error as small as the upper-bound exponent indicates.

The form of Eqs. 4 lends itself to a presentation of E vs P for fixed R. This presentation is shown in Fig. XIV-4. The regions are labeled the same as those in
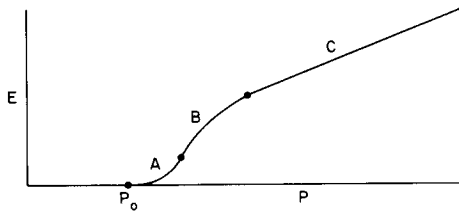
Fig. XIV-4. $E_U(P)$ for fixed R.

Fig. XIV-3. The value P_o is the minimum power needed for reliable transmission at the given rate; and for any power less than P_o, the exponent is zero. In region A we have

$$\frac{\partial E}{\partial P} = \frac{\rho}{2N_b(1+\rho)}.$$

At $\rho = 1$ we go into region B, and here ρ remains constant at 1 but N_b increases, thereby increasing the bandwidth. In this region

$$\frac{\partial E}{\partial P} = \frac{1}{4N_b}.$$

In region C we again find that N_b remains fixed, as in region A, except now at a larger value. We have

$$\frac{\partial E}{\partial P} = \frac{1}{4N_b}$$

which is just a straight line.

This curve gives an attainable exponent for given rate and power. According to the derivation of this bound the power used in each component channel is

$$(1+\rho) \frac{\dfrac{N_b - N_n}{1 - \dfrac{\rho}{1+\rho} \dfrac{N_n}{N_b}}}{},$$

and the total energy used for each block transmission must lie within a shell

$$N(P_{av} - \delta) \leqslant S \leqslant NP_{av}.$$

2. Upper Bound on E(R)

An upper bound on the exponent can be obtained by a sphere-packing type of arguement. This bound has the same form as the lower-bound exponent in region A, Eq. 4, except that now ρ can take on any positive value. A typical curve is shown in Fig. XIV-5.
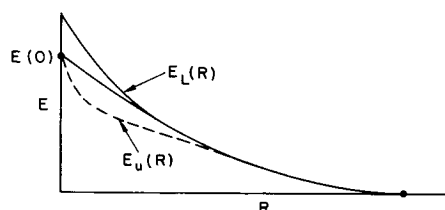
Fig. XIV-5. $E_L(P)$ for fixed P.

There are two small reductions that can be made in this bound. First, the value of E(0) for the lower-bound exponent is $\dfrac{P}{4N_{min}}$ (see Eq. 5), which Shannon[3] has shown is also the upper bound of the white Gaussian channel exponent at zero rate. The upper bound applies to the model also, since it will certainly be inferior to the white channel which has all of its noise power equal to the minimum of the noises in the model. Once the E(0) is reduced, one can produce a straight-line bound through E(0) tangent to the old upper bound by Shannon and Gallager's[2] technique of breaking the channel up into two parts and looking at the best list-decoding bound on one part and the zero-rate bound on the other part. Then the probability of error can be shown to be greater than one-fourth the product of the probability of error for each of these steps. Figure XIV-6 shows the sphere-packing and the zero-rate bounds (not the tangent bound) presented as E vs P. Curves A are the sphere-packing bounds for several rates and have slopes $\dfrac{\rho}{2N_b(1+\rho)}$;
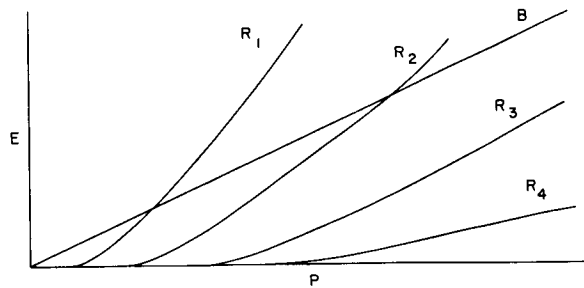
Fig. XIV-6. $E_L(P)$ for several fixed values of R.

thus for large ρ their slopes approach $\frac{1}{2N_b}$. Curve B has slope $\frac{1}{4N_{min}}$ and is independent of rate. It merely states that, no matter what the rate, the exponent is less than the zero-rate exponent. If the rate is large enough, curve A is always below curve B, and thus the tighter bound. The effect of the tangent line is not shown in Fig. XIV-6, since its effect can only be found graphically once the E vs R curves are plotted for all values of P. It is known that this tangent line has no effect for sufficiently large R.

In the region where the upper and lower bounds agree, between R_{crit} and capacity, one can make definite statements about the nature of the optimum signals. It has been found that the signals must lie entirely within those channels, or that part of the spectrum, where the noise power is below the threshold, N_b, and that N_b is determined by the relation

$$R^* = \frac{1}{2} \sum_{N_n \leq N_b} \ln \frac{N_b}{N_n}.$$

In the other regions where the two bounds do not agree, one cannot say anything about the signals with certainty, except that the average power required for a given rate and exponent is less than that given by Fig. XIV-4 and greater than that given in Fig. XIV-6.

3. Outline of Proofs

a. Lower Bound on E(R)

We obtain an upper bound to the probability of error by the standard random-coding argument.[1] The only difference from the standard procedure in this particular case lies in defining the ensemble from which the random code words are chosen. A code word consists of a set of x_n that are to be transmitted through the component channels in the model. The ensemble of codes is defined by taking each x_n in each code word from a Gaussian distribution with zero mean and variance

$$\overline{x}_n^2 = (1+\rho) \frac{N_b - N_n}{1 - \dfrac{\rho}{1+\rho}\dfrac{N_n}{N_b}}; \quad N_n \leqslant N_b$$

$$= 0; \quad N_n > N_b$$

and then rejecting all except those code words for which

$$NP - \delta < \sum_n x_n^2 \leqslant NP; \quad \delta > 0.$$

In other words, the energy in each code word is required to be in a small shell. One then writes, following Gallager,[1]

$$P_e \leqslant e^{-N[E_o(\rho)-\rho R]},$$

where

$$E_o(\rho) = -\ln \int \left[\int p(\underline{x})\, p(\underline{y}/\underline{x})^{\frac{1}{1+\rho}} \, d\underline{x} \right]^{1+\rho} d\underline{y}, \quad 0 \leqslant \rho \leqslant 1.$$

In this equation, $p(\underline{y}/\underline{x})$ is known to be a product of Gaussian distributions. $p(\underline{x})$ is given above but can be bounded for any $r > 0$ by

$$p(\underline{x}) \leqslant q^{-1} e^{r \sum_n x_n^2 - rNP + r\delta} \frac{e^{-\sum_n \frac{x_n^2}{2\sigma_n^2}}}{\prod_n \sqrt{2\pi\sigma_n^2}},$$

where

$$\sigma_n^2 = (1+\rho) \frac{N_b - N_n}{1 - \dfrac{\rho}{1+\rho}\dfrac{N_n}{N_b}}; \quad N_n \leqslant N_b$$

$$r = \frac{\rho}{2N_b(1+\rho)^2}$$

$$q = \int_{NP-\delta \leqslant \sum_n x_n^2 \leqslant NP} \frac{e^{-\sum_n \frac{x_n^2}{2\sigma_n^2}}}{\prod_n \sqrt{2\pi\sigma_n^2}} d\underline{x}.$$

When

$$NP = (1+\rho) \sum_n \frac{N_b - N_n}{1 - \dfrac{\rho}{1+\rho} \dfrac{N_n}{N_b}},$$

q only decreases algebraically with N and consequently does not affect the exponent. Evaluating the expression for $E_o(\rho)$ and maximizing over ρ, we obtain the expressions for R and E. The expurgated bound is found by the same method. Now the variances used are

$$\sigma_n^2 = 4\rho \frac{N_b - N_n}{2 - \dfrac{N_n}{N_b}}; \quad N_n \leq N_b$$

and, following Gallager's expurgated bound,

$$P_e \leq 4\rho e^{-N[E_o(\rho) - \rho R]},$$

where now

$$E_o(\rho) = -\ln \left[\int_{x'} \int_{x} p(x)\, p(\underline{x}') \left(\int_{\underline{y}} \sqrt{p(\underline{y}/\underline{x})\, p(\underline{y}/\underline{x}')}\; d\underline{y} \right)^{1/\rho} d\underline{x}\,d\underline{x}' \right]^{\rho}, \quad \rho \geq 1.$$

This can be evaluated in the same manner, and the expurgated relations can be obtained.

b. Upper Bound on E(R)

The lower bound on the probability of error is obtained by using a method of Gallager,[2] based on the Chebychev inequality. The theorem used here states: Define $\mu_m(s)$ as

$$\mu_m(s) = \ln \int_{\underline{y}} f(\underline{y})^s\, p(\underline{y}/\underline{x}_m)^{1-s}\; d\underline{y},$$

where $f(\underline{y})$ is an arbitrary probability density on $\underline{y}$. Then if

$$\int_{Y_m} f(\underline{y})\; d\underline{y} \leq \frac{1}{4} \exp[\mu_m(s) + (1-s)\mu_m'(s) - (1-s)\sqrt{2\mu_m''(s)}\,],$$

where Y_m is that set of output sequences decoded as m, it follows that

$$P_{em} \geq \frac{1}{4} \exp[\mu_m(s) - s\mu_m'(s) - s\sqrt{2\mu_m''(s)}\,]. \tag{6}$$

Since the sets Y_m are disjoint, $\int_{Y_m} f(\underline{y}) \, d\underline{y}$ cannot be large for all transmitted signals and, in fact, there must be one signal with $\int_{Y_m} f(\underline{y}) \, d\underline{y} \leqslant \frac{1}{M} = e^{-NR}$. Thus when s is chosen so that

$$NR \geqslant -\mu_m(s) - (1-s)\mu_m'(s) + (1-s)\sqrt{2\mu''(s)} + \ln 4, \tag{7}$$

Eq. 6 gives us a lower bound on P_{em}.

We choose $f(\underline{y})$ to be

$$f(\underline{y}) = \prod_n \frac{e^{-\dfrac{y_n^2}{2\sigma_n^2}}}{\sqrt{2\pi\sigma_n^2}},$$

where

$$\sigma_n^2 \begin{cases} = N_n & \text{if } N_n \geqslant N_b \\ \\ = \dfrac{N_b - s_1 N_n}{1 - s_1} & \text{if } N_n < N_b \end{cases}.$$

We shall set s_1 equal to s, but μ_m' and μ_m'' are understood to be the partial derivatives of μ_m with $f(\underline{y})$ fixed. If we set

$$R^* = \frac{1}{2} \sum_{N_n \leqslant N_b} \ln \frac{N_b}{N_n},$$

and then select s to meet Eq. 7 it turns out that the exponential behavior of the lower bound on P_e is the same as the upper bound. One point that needs to be enlarged upon is that $\mu_m(s)$ depends on the m for which Eq. 6 is satisfied; then it depends on $f(\underline{y})$, which in turn depends on s. If one is to choose s to meet Eq. 7, it looks as if an endless circle of dependencies will arise. It turns out that the $\sqrt{2\mu''(s)}$ becomes negligible for large block length, and that the $-\mu_n(s) - (1-s)\mu_m'(s)$ depends on m only through three sums:

$$\sum_{N_n \leqslant N_b} x_{mn}^2,$$

$$\sum_{N_n \leqslant N_b} x_{mn}^2 N_n,$$

$$\sum_{N_n > N_b} x^2_{mn}/N_n.$$

We therefore restrict ourselves to a small fraction of the m. First consider only those m for which

$$\sum_n x^2_{mn} \leq \frac{NP}{1-a};$$

this will be at least aM of the signals. Now the three sums are bounded and can be subdivided into a finite number of intervals, each of length $N\delta$. There must be some triplet of intervals which contains at least $\dfrac{\delta^3(1-a)^3}{\rho^3} aM$ of the input signals. We consider only this set, and note that reducing the set of input symbols by a fixed fraction only reduces the rate by

$$\frac{-\ln \dfrac{\delta^3(1-a)^3}{\rho^3} a}{N}$$

which approaches zero for large N. Once one knows that Eq. 7 can be achieved within $N\delta$ of equality, substitution in Eq. 6 gives the sphere-packing exponent for the probability of error.

P. M. Ebert

References

1. R. G. Gallager, A simple derivation of the coding theorem and some applications (to be published in IEEE Trans., PTGIT).

2. C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, Lower bounds to error probability for coding on discrete memoryless channel (unpublished).

3. C. E. Shannon, Probability of error for optimum codes in a Gaussian channel, Bell System Tech. J. 38, 611 (1959).

Prof. K. N. Stevens	N. Benhaim	J. S. Perkell
Prof. M. Halle	T. H. Crystal	Eleanor C. River
Prof. J. B. Dennis	H. J. Hebert	A. W. Slawson
Prof. J. M. Heinz	W. L. Henke	R. S. Tomlinson
Dr. Paula Menyuk	Y. Kato	J. T. Williams

A. DETECTION OF DSB SIGNALS OCCUPYING THE SAME RF SPECTRUM

1. Introduction

In a recent paper,[1] Bridges and Zalewski have discussed an approximate procedure for reducing the interference of one DSB signal upon another when the signals have overlapping spectra. They compared their scheme with various procedures for taking advantage of bandwidth – SNR tradeoffs – and concluded that it could be used for optimizing bandwidth utilization through the addition and later separation of many DSB signals. Contrary to the results and conclusions of Bridges and Zalewski, we shall show that since DSB is a linear modulation procedure that creates two shifted versions of the modulating signal spectrum, two DSB signals with overlapping spectrum may be perfectly separated by a finite, linear demodulation procedure. Because of the linearity of the modulation and demodulation process, there is no justification for comparison with nonlinear schemes nor for extending this procedure for more signals. The procedure could have important applications for protection of one DSB or AM signal against intentional or chance interference by another.

The ability to separate two overlapping DSB signals is recognized for the singular case of quadrature modulation, when the two signals have the same carrier frequency but one is 90° out of phase with the other. In this case orthogonal detection is used to remove one signal and recover the other. This same technique may be used, even if the carriers are not 90° out of phase, as long as there is some phase difference.

2. Iterative Demodulation Process

The signal to be demodulated r(t) is the sum of two double sideband signals having baseband signals p(t) and q(t), respectively,

$$r(t) = p(t) \cos \omega_p t + q(t) \cos \omega_q t, \tag{1}$$

where $\Delta = \omega_p - \omega_q$ is less than the bandwidth W of q(t), the signal we wish to recover. The phases of the signals have been chosen for mathematical convenience but this imposes

no limitations on the demodulator, except the possible need for some phase shifters. The carriers are assumed to be available at the receiver.

Following Bridges and Zalewski,[1] the signal that is due to p(t) is removed by orthogonal detection, that is, multiplication by $2 \sin \omega_p t$ and lowpass filtering. What remains is

$$r^{(1)}(t) = q(t) \sin \Delta t, \tag{2}$$

the desired signal multiplied by a low-frequency sine wave. The signal q(t) could be recovered, as attempted by Bridges and Zalewski,[1] through multiplication by the reciprocal of $\sin \Delta t$ if there were no zero crossings. The same results may be achieved with a finite process through repeated orthogonal detection:

$$r^{(2)}(t) = r^{(1)}(t) \cdot \cos \Delta t = q(t) \sin 2\Delta t. \tag{3}$$

By this operation, the frequency of the sine wave has been doubled. Repeating the process N times gives

$$r^{(N)}(t) = r^{(1)}(t) \, 2^N \prod_{i=0}^{N-1} \cos (2^i \Delta t) = q(t) \sin (2^N \Delta t). \tag{4}$$

The purpose of repeating this operation is to make the frequency of the sinusoid multiplying q(t) greater than the bandwidth W of q(t). This will be achieved when

$$2^N \Delta \geq W \tag{5}$$

or when

$$N \geq \log_2 \frac{W}{\Delta}. \tag{6}$$

There is no upper limit on N. For N as given above, q(t) may be recovered from $r^{(N)}(t)$ by product demodulation and lowpass filtering.

$$q(t) = \text{L. P. F.} \left\{ r^{(N)}(t) \sin (2^N \Delta t) \right\}$$

$$= \text{L. P. F.} \left\{ q(t) \sin^2 (2^N \Delta t) \right\}. \tag{7}$$

The demodulation process for deriving q(t) from $r^{(1)}(t)$ is the multiplication by the finite product D(t), summarized as follows:

$$D(t) = 2^N \prod_{i=1}^{N} \cos (2^{i-1} \Delta t) \sin (2^N \Delta t), \tag{8}$$

where $N \geq \log_2 \dfrac{W}{\Delta}$.

3. Spectrum and Closed Form of D(t)

The spectrum for D(t) may be found by convolving the spectral lines for each of the factors of the product. The first few and the final application of this procedure are shown in Fig. XV-1. The first N spectra come from the repeated multiplication by cosines, and the $(N+1)^{th}$ from the multiplication by a sine (see Eq. 8). From Fig. XV-1 it can be seen that D(t) contains all of the odd harmonics of Δ from the first to the $(2^{N+1}-1)^{th}$. All components have the same phase, so that D(t) may be written as a Fourier sine series.

$$D(t) = \sum_{k=1}^{2^N} \cdot \sin (2k-1)\Delta t \tag{9}$$

By expressing D(t) as two exponential Fourier series, one each for positive and negative frequencies, and then using the closed form of a geometric progression, we find

$$D(t) = \frac{\sin^2 \Delta Kt}{\sin \Delta t}, \tag{10}$$

where $K = 2^N \geqslant \frac{W}{\Delta}$. Because D(t) is always finite, it may be synthesized and used in a product demodulator. Note that there are lower limits but no upper limits on N and K.

The limit of (10) for $k = \infty$ is not obvious, but if one uses the procedure outlined above for (9), on the following equation

$$D_\infty(t) = \sum_{k=1}^{\infty} \sin (2k-1)\Delta t \tag{11}$$

the result is

$$D_\infty(t) = \frac{1}{\sin \Delta t}. \tag{12}$$

Comparison with (2) shows that multiplying $r^1(t)$ by $D_\infty(t)$ would give q(t), thereby demonstrating that the limiting case of D(t) is the same multiplicand used by Bridges and Zalewski.[1]

4. Noise Performance of the Demodulator

The noise performance of the demodulator is difficult to analyze because the demodulation destroys any assumed stationariness of the input-noise signal. This can be seen from the following development. If r(t) in (1) has bandlimited white noise added to it, $r^{(1)}(t)$ obtained from the first stage of demodulation will similarly have lowpass white noise added to it. If the noise added to $r^{(1)}(t)$ is denoted by N(t), the noise output of the demodulator will be
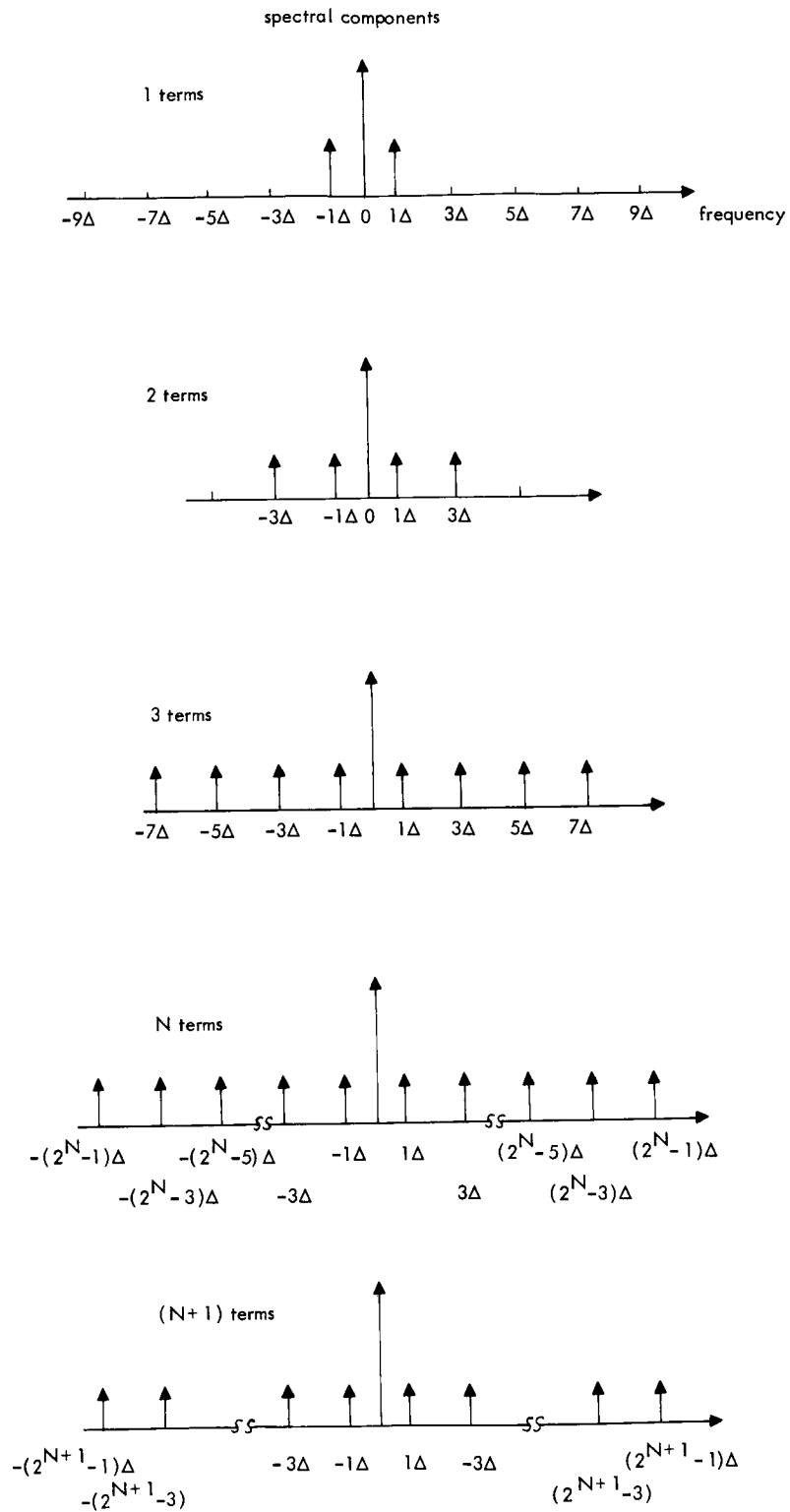
spectral components

1 terms

-9Δ   -7Δ   -5Δ   -3Δ -1Δ 0 1Δ   3Δ   5Δ   7Δ   9Δ   frequency

2 terms

-3Δ   -1Δ 0 1Δ   3Δ

3 terms

-7Δ   -5Δ   -3Δ   -1Δ   1Δ   3Δ   5Δ   7Δ

N terms

$-(2^N-1)\Delta$ $-(2^N-5)\Delta$ -1Δ   1Δ $(2^N-5)\Delta$ $(2^N-1)\Delta$

$-(2^N-3)\Delta$ -3Δ   3Δ $(2^N-3)\Delta$

(N+1) terms

$-(2^{N+1}-1)\Delta$ -3Δ -1Δ   1Δ   -3Δ $(2^{N+1}-1)\Delta$

$-(2^{N+1}-3)$ $(2^{N+1}-3)$

Fig. XV-1. Demodulator spectra.

$$n_o(t) = D(t) \cdot n(t).$$ (13)

One statistic which may be computed is the noise power, $\langle n_o(t)^2 \rangle$ which, since $n(t)$ is independent of $D(t)$, may be expressed as

$$N_o = \langle n_o(t)^2 \rangle = \overline{D^2(t)} \cdot \langle n^2(t) \rangle.$$ (14)

The last factor of (14) is the input noise power, N_i. Dividing both sides of the equation by this factor gives the ratio of output noise power to input noise power:

$$\frac{N_o}{N_i} = \overline{D^2(t)}.$$ (15)

This ratio is equal to the power of $D^2(t)$ and may be computed from the power of the individual components in (9):

$$\frac{N_o}{N_i} = \sum_{k=1}^{2^N} \frac{1}{2} = 2^{N-1}.$$ (16)

This indicates that the noise power increases exponentially with N and that the minimum N consistent with (6) would give the smallest output noise power. Combining this result with that shown in Eq. 5 gives

$$\frac{N_o}{N_i} \geq \frac{1}{2} \frac{W}{\Delta}.$$ (17)

Because (16) does not take into account the lowpass filtering which follows the multiplication by $D(t)$, (17) is probably a good estimate of over-all demodulator performance for all values of K above the minimum.

5. Conclusion

Our discussion has shown that since a pair of overlapping DSB or AM signals (1) may be created by a linear of modulation process and (2) has an RF bandwidth that is at least as great as the sum of the baseband bandwidths, the combined signal may be demodulated and the two modulating signals recovered exactly. The demodulation process does, however, increase the noise power and make the noise probability distribution a periodic function of time.

T. H. Crystal

References

1. J. E. Bridges and R. A. Zalewski, Orthogonal detection to reduce common channel interference, Proc. IEEE 52, 1022-1028 (1964).

B. CUES USED IN SPEECH PERCEPTION AND PRODUCTION BY CHILDREN

In the study of language acquisition and development we would like to discover the particular cues in the language environment which are used by children to determine the linguistically distinct elements, and we would like to know the developmental course of this process. The process occurs from babbling to sentence formation within a short span of time. It seems reasonable to suppose that children have, at an early age, learned to classify speech sounds into grammatically meaningful elements. Also, it seems evident that there are particular cues that are used by children to discover linguistically distinct elements. Some experiments have been undertaken to explore some questions about the phonological component of the grammar which seem relevant to these hypotheses.

Experiment 1

Peterson and Barney[1] found in their experiment concerning vowel identification that children's productions were more difficult to identify than adult productions, and that some vowels are better understood than others. Characteristic shifts in vowel identification were found. They also hypothesized that language experience influences both the production and perception of sounds.

We have undertaken an experiment to explore the effects of the following factors on vowel perception and production: (i) immediate imitation of a model on vowel production by children: (ii) correlations of age of child, sex of child, and sex of adult presenting the stimulus materials on vowel production by children, word context versus nonsense syllable context on vowel production of both children and adults; and (iii) the age of the producer and effect of context on the identification of vowels.

The subjects were a boy and girl, both aged 4, and a boy and girl, aged 10 and 9, respectively. The list of words and the nonsense syllables that were used are shown in Table XV-1.

Table XV-1. Stimulus materials in Experiment 1.

Words	Nonsense Syllables
bid	bɛv
bored	pɪf**
bard	buv
bit*	bɑrv
bead	bʊv
spit**	sprɪf**
bird	bæv
bed	bɪv
beat*	bɔrv
booed	biv
bud	pɪf*
bad	bɜv
pit**	

*Included to examine vowel lengthening.
**Included to examine aspiration.

Each item on the lists was presented to each child for immediate repetition by an adult female, then by an adult male. The order of presentation in each list was changed in each presentation. Presentations and responses were recorded from an anechoic chamber. The tapes are now undergoing spectrographic analysis and will then be presented to children and adults for identification.

Experiment 2

An experiment has been designed to explore the questions of whether or not children can learn to distinguish equally well between members of a set of nongrammatical nonsense syllables and a set of grammatical nonsense syllables, and whether or not this capacity changes with age. The nongrammatical set is composed of initial consonant clusters that are not permissable in English, and the grammatical set is composed of initial clusters that are. Examples are given in Table XV-2.

Table XV-2. Sample stimulus materials in Experiment 2.

Nongrammatical	Grammatical
gz æ k	gl æ k
pfʌm	klʌm
dlɛv	drɛv
srut	sput
dlɛd	krɛd
srɪk	blɪk
pfudʒ	prudʒ
gzos	glos

The task has three parts: (i) to learn to associate a nonsense syllable with a colored circle, each circle having the same color and having the name of one member of the set, and all circles presented simultaneously in a horizontal array; (ii) to immediately repeat all members of a different pair of sets after the experimenter; and (iii) to repeat each member of a set, one by one, after the experimenter. The first part of the task may give us information about differences, if any, in the perception and storing of nongrammatical and grammatical speech material, the second, differences in perecption, storing and reproduction of materials, and the third, differences in capacity to articulate the materials.

In a preliminary test of the stimulus materials, with two subjects, aged 5 and 9, used, it was found that only 4 members in a set for the first part of the task was too simple for both children, independently of the kind of material presented. Learning occurred after one or two trials. The second part of the task was very difficult, even with only 4 members in a set. Perfect repetition of the nongrammatical set was not accomplished after many trials by either child and long after exhaustion and disinterest had occurred. A sample study will be undertaken with small groups of children, aged from 3 to 7 years,

with an expanded set for the first part of the task. Unlike the preliminary test, the two kinds of materials will be presented one week apart to counteract both learning effects and fatigue.

Further Research

An analysis of some available data on the perception and production of consonants by both adults and children has been undertaken. We have postulated that by describing the behavior observed in terms of a distinctive feature matrix[2] more information could be derived about cues used in the perception and production of consonants by children. The features looked at were: grave, diffuse, strident, nasal, continuant, and voice. We felt that this kind of analysis might possibly lead to productive research with children.

The data looked at were confusion in the perception of consonants by adults,[3,4] the consonant substitutions of children with functionally deviant speech[5], speech sounds which have been found to be most frequently defective in analyses of children's and adults' speech[6], and the acquisition and proportion of usage of consonants from age 3 months to 8 years[7,8] and adults.[9]

There has been very little research on the perception of consonants by children, and practically none in the age period before school. Also, it is difficult to determine, whether the consonants described as being acquired in developmental studies are in the ear of the hearer or the mouth of the child, since the data were gathered by phonetic transcriptions of either recorded or nonrecorded speech, rather than by spectrographic analysis. Also, a distinction must be made between productive use of consonants (as in morpheme formation) and production of consonants.

In this preliminary examination, we found that in general the features voicing and nasality were largely resistant to perceptual and productive confusion, and weighed heavily in the proportion of usage of consonants from age 3 months to 4 years. The importance of place (gravity and diffuseness) in proportion of usage differed over the age range of 3 months to adulthood. Gravity weighed more heavily during the early age periods of from 3 months to 30 months, and again of from 2 years to 8 years; then, diffuseness assumed a greater importance. In both the consonant substitutions by children and the perceptual confusions of adults, however, place was the feature that was least maintained. Consonants with the features stridency and continuancy are used proportionately less by both children and adults. They are consonants which are acquired last by children and they compose the list of those consonants most frequently found defective in the analyses that have been carried out.

As a possibility for future research, taking into account the results of the completed analysis of the data, one could look at the question of rank ordering of features in terms of ease of discrimination by young children, and even, possibly, infants.

Paula Menyuk

References

1. G. Peterson and H. L. Barney, Control methods used in a study of vowels, J. Acoust. Soc. Am. 24, 175-184 (1952).

2. R. Jakobson, M. Halle, and C. G. M. Fant, Preliminaries to speech analysis, Technical Report 13, Acoustics Laboratory, M. I. T., June 1955.

3. G. A. Miller and P. E. Nicely, Analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Am. 27, 338-352 (1955).

4. W. Wicklegren, Psychology Department, M. I. T., Personal communication, 1964.

5. Paula Menyuk, Comparison of grammar of children with functionally deviant and normal speech. J. Speech Hearing Res. 7, 109-121 (1964).

6. M. H. Powers, Functional Disorders of Articulation, Handbook of Speech Pathology, L. E. Travis, (ed.)(Appleton, Century, Crofts, New York, 1957), pp. 707-769.

7. O. C. Irwin, Infant speech: Consonant sounds according to manner of articulation, J. Speech and Hearing Disord. 12, 397-401 (1947).

8. D. McCarthy, Research in language development: Retrospect and prospect, Monograph, Society for Research on Child Development, Vol. 24, pp. 3-25, 1959.

9. P. B. Denes, On the statistics of spoken English, J. Acoust. Soc. Am. 35, 892-905 (1963).

XVI. MECHANICAL TRANSLATION[*]

V. H. Yngve	Ursula Funke	Elizabeth Landers
Carol M. Bosche	Elfriede Goody	W. A. Lea
Elinor K. Charney	T. R. Hofmann	Ellie Reed Lewis
J. L. Darlington	L. Jenkins	F. Liu
J. M. Dolan	Muriel S. Kannel	B. K. T'sou
R. S. Fabry	E. S. Klima	R. J. Zatorski[†]

A. REVISED PROOF PROCEDURE PROGRAM

In Quarterly Progress Report No. 76 (pages 260-270), we described a COMIT program that proves logical arguments expressed in a restricted form of ordinary English. Two examples were given of arguments that the logical evaluation part of the program was unable to prove, even though it was verified by hand computation that the translations into logical symbolism were correct and that the arguments were valid. A more efficient logical proof procedure program has since been devised.[1] It not only proves arguments that the previous program was unable to prove, but it also substantially reduces the computation time (in some cases by a factor of 10) for some arguments and theorems that the previous program was able to prove. While the previous program was based to a large extent on the Davis-Putnam proof procedure algorithm,[2] the new program is an amalgam of the "one-literal clause rule" of Davis-Putnam and the "matching algorithm" of J. R. Guard.[3] Taking its point of departure from the Davis-Putnam algorithm, the program negates an input formula and puts the negated formula into prenex normal form and the matrix into conjunctive normal form. At this point, instead of manufacturing generations of quantifier-free substitution instances (QFSI) and testing for consistency at certain specified points after the manner of Davis-Putnam, the program attempts to produce a proof on the basis of the "one-literal clauses" (i. e., conjuncts containing no disjunction operators) in the matrix. By using Guard's matching algorithm, which enables one to test whether two clauses could generate any common QFSI, the program tests the one-literal clauses in the matrix for consistency. If two such clauses match negatively ("N-match"), i. e., if the two clauses could generate contradictory QFSI, then the matrix is inconsistent and the original formula is valid. If the one-literal clauses are consistent, then the program attempts to generate additional one-literal clauses by N-matching the existing one-literal clauses against the polyliteral clauses. For example, the one-literal clause

$$F(y, P(x, y))$$

— or FYPXY, in COMIT notation — N-matches the polyliteral clause

$$not-F(x, y) \lor G(y, x)$$

— or NFXY + GYX, in COMIT notation — to produce the one-literal clause

$$G(P(x,y),y)$$

— or GPXYY, in COMIT notation. Whenever the N-matching procedure produces a new one-literal clause, it is tested for consistency against the existing one-literal clauses. If it is inconsistent with the others, it means that the original formula is valid, and the run is terminated. If one or more polyliteral clauses are produced by this N-matching procedure, they will have one less term than the original clause, and may therefore produce one-literal clauses during a further N-matching. If all possible N-matchings fail to produce a contradictory set of one-literal clauses, or if there were no one-literal clauses to start with, then the program reverts to the older Davis-Putnam method of generating QFSI. The program tests the QFSI for inconsistency after each generation, using an algorithm based on the "one-literal clause rule," which is the most efficient of the three Davis-Putnam rules for propositional logic, and which may be stated as follows: Given a formula C in conjunctive normal form, and a one-literal clause P of C, all conjuncts containing P are deleted from C, and all single occurrences of not-P are deleted from C, thereby producing a reduced formula C' that is consistent if and only if C is consistent.

Since the deletion of single occurrences of not-P may produce new one-literal clauses, the one-literal clauses must be tested for consistency after each application of the rule. The condition of inconsistency is the occurrence of two contradictory one-literal clauses; and the condition of consistency is the deletion of the entire formula C. If the one-literal clause rule fails to prove C either consistent or inconsistent, then C is split into two branches by assuming a term P first true and then false, and making appropriate cancellations according to a method described by Quine.[4] The first branch is formed by deleting all conjuncts containing P and all single occurrences of not-P, and the second branch is formed by deleting all conjuncts containing not-P and all single occurrences of P. The second branch is stored at the front of a shelf, and the first branch remains in the workspace where an attempt is made to apply the one-literal clause rule to it. If this attempt fails, the formula in the workspace is split again in the same way, the first branch remaining in the workspace and the second branch being stored at the front of the shelf. This procedure continues until a branch is obtained to which the one-literal clause rule can be applied. The consistency test terminates when one branch is proved consistent (i. e., is entirely deleted), in which case the entire formula is consistent, or when all branches are proved inconsistent (i. e., contain contradictory one-literal clauses), in which case the entire formula is inconsistent.

<div align="right">J. L. Darlington</div>

References

1. The revised program is described in greater detail in a memorandum, "A proof procedure program based on the 'one-literal clause rule' and the 'matching algorithm'," Mechanical Translation Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, March 1965.

2. M. Davis and H. Putnam, A computing procedure for quantification theory, J. ACM 7, 201-215 (1960).

3. J. R. Guard, Automated Logic for Semi-automated Mathematics, Scientific Report No. 1, Contract No. AF19(628)-3250, Project No. 5632, Task No. 563205, Applied Logic Corporation, Princeton, New Jersey, March 30, 1964.

4. W. V. Quine, Methods of Logic (Henry Holt Company, New York, 1959), pp. 26-27.

B. CONSTITUENT STRUCTURE AND WORD-ORDER RULES FOR GERMAN

The following rules represent a segment of the grammar of contemporary standard German. This miniature grammar is directed especially toward the solution of three classical problems in German syntax: (a) the order of nominals (subject and objects) and adverbials with respect to one another and to the finite verb; (b) the way in which independent declarative sentences differ, in particular, from subordinate clauses with respect to this order; and (c) the occurrence of positional es, as in "es kommen viele Leute."

The aspect of the constituent structure which is original is the treatment of the subject nominal (Nom_0) as the first among equals in a series of verbal specifiers which includes the object nominals (including the direct object Nom_1 and the indirect object Nom_2), as well as the adverbial adjuncts (among others, adverbials of location Loc and of manner Mod). All of these specifiers of the verb are immediately dominated by a single common constituent labeled Sp.

The constituent structure presented below differs from those assumed in previous studies in the transformational syntax of German[1] by having an empty constituent (Topic) into which some immediate constituent of Sp may, or under certain conditions must, be incorporated by a transformational rule of Topicalization.

The constituent structure resembles that assumed by Matthews and Rogovin, by Bach, and by Bierwisch, in that the verbal complex originates at the end of the sentence, with finite verb last. A transformational rule of Finite verb placement correctly locates the finite verb in second position (i. e. , after the Topic or after the interrogative marker in independent sentences).

The symbols used here have the following interpretations: S sentential clause, W interrogative marker, Kn subordinating conjunction, R relative conjunction, Nom nominal (with subscripts $_0, _1, _2$ referring to nominative, accusative, and dative, respectively),

Loc adverbial of location, Mod adverbial of manner, Adj predicate adjective, Sp speci-
fier of the verb, V principal verb, sep separable prefix, Aux verbal auxiliary, M modal
verb, T finite tense and person marker, VP verb phrase, # sentence boundary, +pp past
participial form, Pro personal pronoun, Reflex reflexive pronoun, inf infinitival form.

Constituent structure:

$$S \longrightarrow \# \text{ Vorfeld } - VP \#$$

$$\text{Vorfeld} \longrightarrow \left\{ \begin{array}{l} W(Kn) \\ (R)Kn \\ Topic \end{array} \right\}$$

$$VP \longrightarrow (Sp) \, V - Aux$$

$$Sp \longrightarrow (Nom_0)(Nom_2)(Nom_1) \left(\left\{ \begin{array}{l} Mod \\ Adj \end{array} \right\} \right) (Loc)$$

$$Aux \longrightarrow \left\{ \begin{array}{l} (Passive)(M)(M) \\ (Passive) \text{ Perfect } (M) \\ (M)(Perfect) \end{array} \right\} \, T$$

$$Topic \longrightarrow es$$

Transformational structure

I.   Topicalization – conditional:

$$\# \quad es[X^1 - sp - \underbrace{X^2]_{Sp} \quad X^3}$$

$$\begin{array}{cccccc}
 & 1 & 2 & 3 & 4 & 5 \\
\Rightarrow & 1 & 4 & 3 & 0 & 5
\end{array}$$

where $[\ldots sp \ldots]_{Sp}$ signifies any constituent immediately dominated by SP.

IIa.   Intraspecifier word order for personal and reflexive pronouns – conditional:

$$\underbrace{\# \, X^1 \Big[X^2}_{} - sp \left\{ \begin{array}{l} Pro \\ Reflex \end{array} \right\} \underbrace{X^3 \Big]_{Sp} \, X^4 \#}_{}$$

$$\begin{array}{cccc}
1 & 2 & 3 & 4 \\
1 & 3 & 2 & 4
\end{array}$$

Obligatory if: $$\underbrace{\# \, X^1 \Big[X^2}_{1} \left\{ \begin{array}{l} Nom_1 \\ Nom_2 \end{array} \right\}_{2} \left\{ \begin{array}{l} Pro \\ Reflex \end{array} \right\}_{3} \underbrace{X^3 \Big]_{Sp} \, X^4 \#}_{4}$$

Optional if: $$\underbrace{\# \, X^1 [X^2}_{1} - \underbrace{Nom_0}_{2} - \underbrace{Reflex}_{3} - \underbrace{X^3]_{Sp} \, X^4 \#}_{4}$$

IIb. Intraspecifier word order for newly introduced versus already mentioned specifiers — conditional:

$$\underbrace{\# \; \bar{X}^1 \; [\bar{X}^2}_{} - sp^1 - sp^2 \; \underbrace{\bar{X}^3 \,]_{Sp} \; \bar{X}^4}_{} \; \#$$

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ \Longrightarrow 1 & 3 & 2 & 4 \end{array}$$

The transformation has least conditions on its applicability when sp^1 is a non-unique, not previously mentioned specifier (e. g., with indefinite article), and sp^2 is an already mentioned specifier (e. g., with definite article). The transformation is nonapplicable if sp^1 is a pronoun and sp^2 is a noun.

III. Placement of the perfect auxiliary of modal verbs — obligatory (A similar optional rule operates on <u>sehen, hoeren</u> in place of <u>M</u>.):

$$\underbrace{\# \; X}_{} \;\; \underbrace{(sep) \; V - M}_{} - \underbrace{Perfect - T}_{} \; \#$$

$$\begin{array}{ccccc} 1 & 0 & 2 & 3 & 4 \\ \Longrightarrow 1 & 3 & 2 & 0 & 4 \end{array}$$

IV. Creation of past participles and finite verb:

$$X \; \begin{Bmatrix} V \\ Perfect \\ Passive \end{Bmatrix} \; \underbrace{\begin{Bmatrix} Perfect \\ Passive \end{Bmatrix} \; X}_{}$$

$$\begin{array}{ccc} 1 & 2 & 3 \\ \Longrightarrow 1 & 2+pp & 3 \end{array}$$

$$X \; \begin{Bmatrix} V \\ M \\ Perfect \\ Passive \end{Bmatrix} \; T - X$$

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ \Longrightarrow 1 & 2+3 & 0 & 4 \end{array}$$

V. Conversion of uninflected verbs into infinitives — obligatory:

$$X \; \begin{Bmatrix} V \\ M \\ Perfect \\ Passive \end{Bmatrix} - X$$

$$\begin{array}{ccc} 1 & 2 & 3 \\ \Longrightarrow 1 & 2+inf & 3 \end{array}$$

where simple concatenation and affixation (indicated by +) are differentiated.

VI.   Finite verb placement:

$$\# \underbrace{\left[\begin{Bmatrix} W \\ Topic \end{Bmatrix}\right]_{Vorfeld}}_{\substack{1 \\ 1}} \quad X^1 \underbrace{\begin{Bmatrix} V \\ M \\ Perfect \\ Passive \end{Bmatrix}}_{\substack{0 \quad 2 \qquad 3 \\ 3 \quad 2 \qquad 0}} + T - \underbrace{X^2 \#}_{\substack{4 \\ 4}}$$
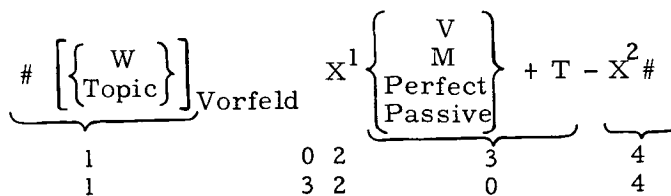
Comments

Among the conditions on Topicalization are the following: (a) if no other constituent immediately dominated by <u>Sp</u> has been topicalized, then a pronoun subject must be topicalized   ("es kommen viele Leute" without topicalization and "viele Leute kommen" with topicalization, but no*"es kommt er," only "er kommt");   (b) under certain conditions, a noun already mentioned in a preceding sentence and thus determined by a definite article or its equivalent must be topicalized ("der Mann kam an" versus *"es kam der Mann an");   (c) certain pronoun objects may not be topicalized (*"sich ereignete ein Unglueck" versus "es ereignete sich ein Unglueck" with positional <u>es</u> and without topicalization, on the one hand, and "ein Unglueck ereignete sich," on the other hand).

The rules for Intraspecifier word order account for the fact that the permissible initial constituents following the interrogative word, if there is one, in questions and following the conjunction in subordinate clauses do not coincide with the constituents that may occur as initial constituent (i.e., as <u>Topic</u>) in independent declarative sentences. Thus, <u>im Zimmer</u> may occur initially as <u>Topic</u> in an independent declarative sentence even if the sentence has a pronoun subject ("im Zimmer spielte er Karten"), but not in a subordinate clause (*"weil im Zimmer er Karten spielte,... "), whereas if the subject is a noun (with an indefinite article, particularly), the permutation is possible even in subordinate clauses ("weil im Zimmer ein Gast Karten spielte,... "). Moreover, in a subordinate clause, a reflexive pronoun may occur initially ("weil sich ein Unglueck ereignete,... "). As we have mentioned, topicalization of a reflexive pronoun is not possible (*"sich ereignete ein Unglueck").

Transformational rules III, IV, and V account for the following facts: (a) in subordinate clauses the finite perfect auxiliary (<u>habe</u>, <u>hat</u>, etc.) forming the periphrastic perfect of a modal verb occurs at the head, rather than at the end, of the verbal complex, i.e., in front of the principal verb and its separable prefix ("weil er den brief hat schreiben wollen,..." versus "weil er den Brief schreiben wollen wird,...," "weil er den Brief geschrieben hat,..." or "weil der Brief geschrieben worden ist,..."); this order holds only if the modal verb is preceded by an infinitive ("weil er den Brief gewollt hat,..."); (b) the modal verb itself, when immediately preceded by an infinitive, assumes

its infinitival rather than its past participial form in this periphrastic perfect construction; this is the case in independent sentences, as well as in subordinate clauses ("er hat den Brief schreiben wollen" versus "er hat den Brief gewollt"); (c) the ungoverned citation form of verbs, as well as this special perfect, is the infinitive. The grammar accounts for these characteristics by a particular interpretation of German syntax. Rule IV (Creation of past participles and finite verb) interprets verbal government as regressive; i. e., the form of the verb is determined by the constituent to its right. Thus, geben assumes the form gab if one particular form of T is to its right, namely third person singular past; geben assumes the form gegeben if either the Passive (werden) or the Perfect (haben or sein) is to the right. By rule V, any verbal stem that has not undergone some process of suffixation (the incorporation of a marker of tense, a past participial marker, or, though not included in the present miniature grammar, a nominalizing or a present participial marker) automatically assumes its infinitival form. The fact that under certain conditions the modal verbs have infinitival form in the perfect is interpreted as a result of the relative ordering of rules. The rule that places the perfect auxiliary at the head of the verbal complex (rule III) precedes the rules whereby the finite or participial form of a verb, if it is to occur in one of those forms, is determined by the constituent that follows it. The result of this ordering of rules is that the modal verb, since it is no longer followed by the auxiliary of the perfect, is therefore not subject to conversion into its past participial form and thus by rule V becomes an infinitive.

The rule of Finite verb placement in independent sentences (VI) is ordered after the rule for the Placement of the perfect auxiliary of modal verbs and has in its domain not only any finite verb, including the auxiliary of the perfect, which occurs at the end of the verbal complex but also the perfect auxiliary relocated by rules III at the head of the verbal complex. In this way the occurrence of the infinitival perfect form of modal verbs in independent sentences is also accounted for (not only "er hat den Brief gewollt" from an underlying $[er]_{Topic}$ den-Brief-woll-Perfect-T, but also "er hat den Brief schreiben wollen" from an intermediate $[er]_{Topic}$ den-Brief-Perfect-T-schreib-woll).

The following fact about German grammar lends support to the assumption (a) that Finite verb placement in independent sentences (rule VI) occurs after the Placement of the perfect auxiliary of modal verbs (rule III) and after the rule for the Creation of past participles (IV), and (b) that rule VI applies also to the relocated finite perfect auxiliary that heads the verbal complex: namely, for the perfect of certain semimodal verbs like hoeren, the perfect auxiliary is found in subordinate clauses both at the end of the verbal complex, in which case the semimodal must be in its past participial form, gehoert, and at the head of the verbal complex, in which case the semimodal must be in its infinitival form ("weil er ihn abfahren gehoert hat,..." and "weil er ihn hat abfahren hoeren,..."). The double source of the finite perfect auxiliary correctly predicts the occurrence of both infinitival and past participial forms in independent

sentences of otherwise identical form ("er hat ihn abfahren gehoert" from $[er]_{Topic}$ ihn-ab-fahr-hoer-Perfect-T and "er hat ihn abfahren hoeren" from an optional intermediate $[er]_{Topic}$ ihn-Perfect-T-ab-fahr-hoer).

E. S. Klima

References

1. G. H. Matthews and Syrell Rogovin, German sentence recognition, Mechanical Translation 5, 114-20 (1958); R. B. Lees, Structural grammars, Mechanical Translation 4, 5-10 (1957; E. Bach, The order of elements in a transformational grammar of German, Language 38, 263-9 (1962); M. Bierwisch, Grammatik des deutschen Verbs, Studia Grammatica II (Berlin, 1963).

XVII. COGNITIVE INFORMATION PROCESSING[*]

Prof. S. J. Mason
Prof. M. Eden
Prof. T. S. Huang
Prof. O. J. Tretiak
Prof. D. E. Troxel
Dr. P. A. Kolers
C. E. Arnold
Ann C. Boyer
J. K. Clemens

R. W. Cornew
J. E. Cunningham
E. S. Davis
R. W. Donaldson
J. K. Dupress
C. L. Fontaine
A. Gabrielian
G. S. Harlem

K. R. Ingham
L. B. Kilham
E. E. Landsman
F. F. Lee
R. A. Murphy
Kathryn F. Rosenthal
F. W. Scoville
N. D. Strahm
Y. Yamaguchi

A. COGNITIVE PROCESSES

1. INTERLINGUAL TRANSFER OF READING SKILL

In previous reports we have shown that a stable order characterizes the ease with which college students can read text that has been transformed geometrically: equal amounts of practice with mathematically equivalent transformations do not yield equivalent levels of performance. Some transformations are considerably more difficult than others.[1] Practice on any transformation, however, facilitates performance on any other; this suggests a generalized habituation to the fact of transformation itself. How generalized that habituation is was studied in the experiment described here.

Ten bilingual subjects, German nationals who had been in the United States for at least nine months, were tested. All were students at the Massachusetts Institute of Technology. Five of these men read 20 pages of English that had been printed in inverted form, and then read 4 pages of German in the same transformation; the other five read 20 pages of German in inverted form, and then 4 pages of English. Also, on the first day, before reading any of the transformed text, and on the fourth day, after all of the transformations had been read, all of the subjects read 1 page of normal English and 1 page of normal German. The time taken by the subjects to read each page was measured with a stop watch. The results are shown in Fig. XVII-1. The speed with which transformed English or German is read increases sharply with practice, from an initial 13 min/page to approximately 4 min/page. Even the latter rate, however, is considerably slower than that for normal text, while normal English (circles) takes a little longer than normal German (triangles). The transfer tests, however, produce asymmetrical results. The subjects trained on 20 pages of inverted English (closed circles) then read four pages of inverted German with no change in the level of performance; but the subjects trained on inverted German (crosses) did not do as well when tested on inverted English.
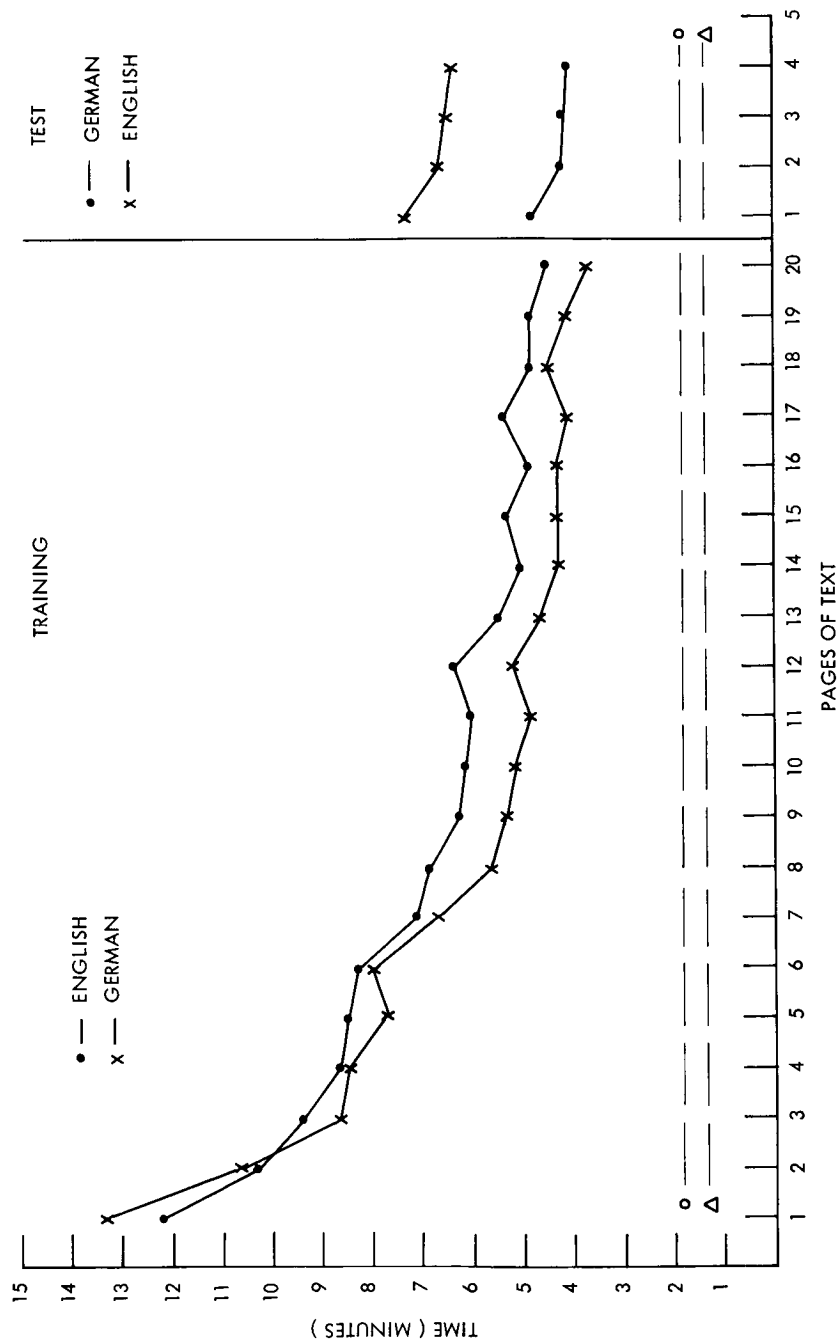
Fig. XVII-1. Results of transfer tests.

324

This curious asymmetry of transfer has an analog in a number of sensori-motor coordinations, for which the general finding is that practicing the less favored organ permits more transfer to the more favored than the reverse direction does; for example, training a right-handed man's left hand on a complex task enables him thenceforward to perform the task with his right hand, but training his right hand does not usually enable him to perform the task with his left.[2] In the present case we find that training in English enabled native speakers of German to transfer their skill without decrement, but training in German yielded some decrement for performance in English.

The more interesting aspect of these results has to do with what is learned when a subject learns how to decode transformed text. If he were learning only to recognize letters that had been transformed, transfer between the languages would be perfect, for the German and English alphabets are almost identical when Roman type is used, the only difference being the use of the diaresis, which does not affect letter shapes. If he were learning the shapes of words, transfer would be relatively poor, since German and English word shapes are somewhat different. The results indicate that the learning cannot be as simple as either of these alternatives would have it.

P. A. Kolers, Ann C. Boyer

References

1. P. A. Kolers, M. Eden, and Ann Boyer, Reading as a perceptual skill, Quarterly Progress Report No. 74, Research Laboratory of Electronics, M.I.T., July 15, 1964, pp. 214-217.

2. R. S. Woodworth, Experimental Psychology (Henry Holt and Company, New York, 1938).

B. PICTURE PROCESSING

1. OPTIMUM SCANNING DIRECTION IN TELEVISION TRANSMISSION

In television transmission, the two-dimensional picture is first scanned to produce a video signal which is then sent through some channel to the receiver. At the receiver, the picture is reconstructed from the video signal by scanning. For any given picture, different scanning methods usually give rise to different video signals and reconstructed pictures. In this report we shall discuss the relative merits of the members of a subclass of scanning methods. We restrict our attention to constant-speed sequential scanning along equidistant parallel lines of slope a (Fig. XVII-2) and try to study the effect of scanning direction on the video signal and the reconstructed picture.

First, we shall find the direction of scanning (that is, the value of a) which yields the minimum-bandwidth video signal, assuming that the two-dimensional Fourier spectrum of the original picture is given. Then we describe some preliminary results
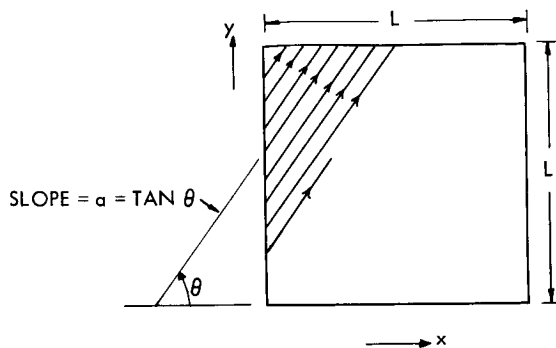
Fig. XVII-2. A subclass of scanning methods.

concerning the subjective effect of scanning direction. Finally, we shall discuss some miscellaneous factors that might affect the choice of a scanning direction.

Minimization of Video Signal Bandwidth

For the sake of simplicity, we shall assume in the following analysis that there is no interlacing. Notice, however, that the addition of interlacing will not change the result of the analysis.

Consider a single picture frame. Let $f(x, y)$ denote the brightness of the picture point (with the average value substracted) as a function of its spatial coordinates (x, y), under the assumption that $f(x, y) = 0$ if (x, y) lies outside the picture. Let $\phi(\tau_1, \tau_2)$ be the autocorrelation function of $f(x, y)$; and $\Phi(u, v)$, the Fourier transform of $\phi(\tau_1, \tau_2)$, that is, the energy spectrum of $f(x, y)$. Let $\phi_a(\tau)$ and $\Phi_a(\omega)$ be the autocorrelation function and the energy spectrum, respectively, of the video signal $f_a(t)$, derived from $f(x, y)$ by scanning along the direction a. The question is: If $\Phi(u, v)$ is given, what value of a will give the minimum bandwidth $\Phi_a(\omega)$? Without loss of generality, we assume that the scanning speed is 1 unit length/unit time. We also assume that the energy of the picture signal is much larger than that of the synchronous and blanking pulses so that the latter can be neglected. Then, we have

$$\phi_a(\tau) \approx \phi(\tau \cos \theta, \ \tau \sin \theta), \tag{1}$$

where $\theta = \tan^{-1} a$, assuming that both the distance between sucessive scanning lines and the width of $\phi(\tau \cos \theta, \ \tau \sin \theta)$ are much smaller than L, the width of the picture. In the case $a \approx 0$ or ∞, $\phi_a(\tau)$ will have peaks at multiples of L, and the right-hand side of Eq. 1 gives only the central peak (at $\tau = 0$); however, the bandwidth of $\Phi_a(\omega)$ is determined mainly by the central peak. In the sequel we shall assume that (1) is an equality.

It follows[1] from Eq. 1 that

$$\Phi_a(\omega) = \frac{1}{2\pi|\cos\theta|} \int_{-\infty}^{\infty} \Phi\left(\frac{\omega}{\cos\theta} - av, v\right) dv. \tag{2}$$

Assuming that $0 \leqslant \theta \leqslant \frac{\pi}{2}$, so that $\cos\theta \geqslant 0$, we have

$$\Phi_a(\omega) = \frac{\sqrt{1+a^2}}{2\pi} \int_{-\infty}^{\infty} \Phi\left(\sqrt{1+a^2}\,\omega - av, v\right) dv. \tag{3}$$

Let us define the bandwidth of $\Phi_a(\omega)$ as

$$B_a \equiv \frac{\int_{-\infty}^{\infty} \Phi_a(\omega)\, d\omega}{\Phi_a(0)}. \tag{4}$$

This definition is reasonable because $\Phi_a(\omega) \geqslant 0$ for all ω, and for most pictures, $\Phi_a(\omega)$ have their maxima at $\omega = 0$. Now

$$\int_{-\infty}^{\infty} \Phi_a(\omega)\, d\omega = 2\pi\phi_a(0) = 2\pi\phi(0,0) \tag{5}$$

is independent of a. So in order to minimize B_a, we have to maximize

$$\Phi_a(0) = \frac{\sqrt{1+a^2}}{2\pi} \int_{-\infty}^{\infty} \Phi(-av,v)\, dv. \tag{6}$$

Hence, we want to choose that value of a which will maximize the right-hand side of Eq. 6. Referring to Fig. XVII-3, we have
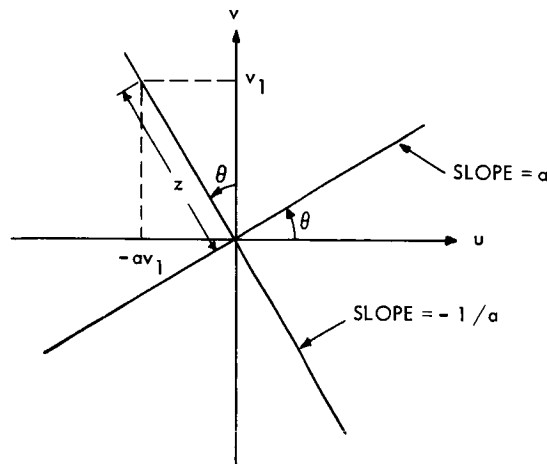
Fig. XVII-3. Pertaining to the interpretation of the right-hand side of Eq. 6.

$$\sqrt{\frac{1+a^2}{2\pi}} \int_{-\infty}^{\infty} \Phi(-av, v) \, dv = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi(-z \sin\theta, z \cos\theta) \, dz. \tag{7}$$

Notice that $\Phi(-z \sin\theta, z \cos\theta)$ is the value of $\Phi(u, v)$ along the straight line $v = -\frac{1}{a}u$, which is perpendicular to the straight line $v = au$. To minimize B_a, therefore, we want to maximize the right-hand side of (7), or equivalently, to maximize

$$W_{-\frac{1}{2}} \equiv \frac{\int_{-\infty}^{\infty} \Phi(-z \sin\theta, z \cos\theta) \, dz}{\Phi(0, 0)} \tag{8}$$

which is defined as the bandwidth of $\Phi(u, v)$ along the direction $-\frac{1}{a}$.

We conclude, therefore, that in order to obtain the video signal of minimum bandwidth, one should scan the original picture along a direction perpendicular to the direction of maximum bandwidth of $\Phi(u, v)$. This result is perhaps not in accord with one's intuition because, intuitively, one might think that to obtain the minimum-bandwidth video signal, one should scan along the direction of minimum bandwidth of $\Phi(u, v)$; this is not the case according to our analysis.

To verify the result of our analysis, we generated some two-dimensional lowpass Gaussian noise with power density spectra (Fig. XVII-4).

$$\Phi(u, v) = \begin{cases} \dfrac{\text{Constant}}{k_1, k_2}, & \text{for } -k_1 \leq u \leq k_1, \text{ and } -k_2 \leq v \leq k_2 \\[2ex] 0, & \text{elsewhere} \end{cases} \tag{9}$$

where k_1 and k_2 are positive real constants. The results of this noise generation (with DC level added) are shown in Fig. XVII-5. According to our analysis, to obtain the minimum-bandwidth video signal, we should scan along the directions $\pm\frac{k_1}{k_2}$. The appearance of the noise does seem to verify our contention.

We note in passing that if the scanning speed and the distance between successive scanning lines (which is assumed to be much smaller than L) are kept constant, then the scanning time per picture frame is independent of the direction of scanning.

Subjective Effect of Scanning Direction

At the ordinary viewing distance (4 or 6 times the picture height), one can see the line structures in the received picture. Do people prefer line structures of a particular orientation to those of other orientations? To try to find an answer to this question, we generated pictures scanned along various directions on a closed-circuit television system. Some of these pictures are shown in Fig. XVII-6.

We showed these pictures to some of our colleagues, and we have listed their

$$\Phi(u,v) = \begin{cases} \dfrac{\text{CONSTANT}}{k_1 k_2}, & \text{IN SHADED REGION} \\[2mm] 0, & \text{ELSEWHERE} \end{cases}$$
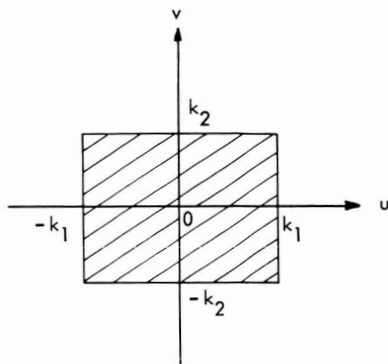
Fig. XVII-4.   Spectrum of two-dimensional lowpass Gaussian noise.
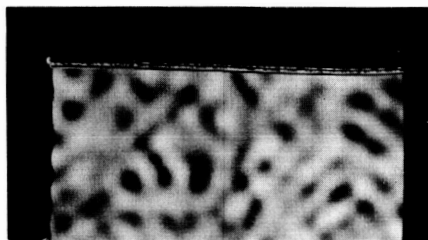
(a)
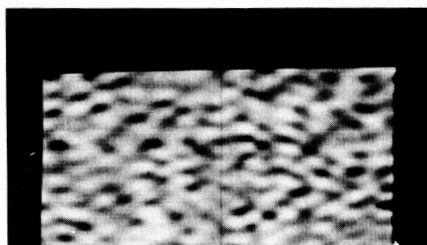
(b)

Fig. XVII-5.   Two-dimensional lowpass Gaussian noise. (a) $k_1/k_2 = 1$.
(b) $k_1/k_2 = \dfrac{1}{2}$.

(a)
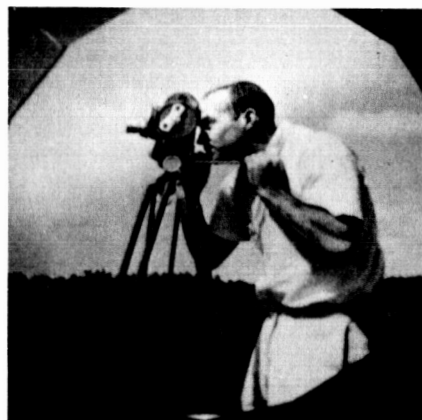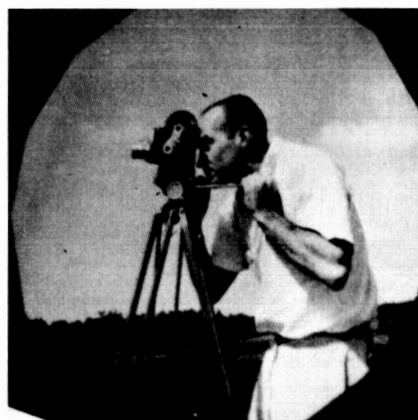
(b)   (c)

Fig. XVII-6.   Picture scanned along various directions.

preferences as follows.

Orders (in the order of decreasing preference):

Subject A: Vertical, horizontal, skew.

Subject B: Horizontal, skew, vertical.

Subject C: Horizontal, vertical, skew.

Subject D: Skew, horizontal, vertical.

The preference, however, was by no means strong.

It is interesting to note that Subject C disliked skew scanning because it seemed to cause anxiety, while Subject D liked skew scanning because it made the picture look "dynamic." Subject B disliked vertical scanning because vertical lines seemed most

visible, and Subject D disliked vertical scanning because the picture seemed ready to fall apart.

Jumping to a tentative conclusion, we might say that the preference is not strong but horizontal scanning seems to have a slight lead.

Other Factors

The pictures mentioned in the preceding section are essentially noiseless. In practice, however, the received picture contains additive random noise and ghosts (caused by multipath). How do the effects of random noise and ghosts depend on scanning direction? Also, how is motion affected by scanning direction? These questions are being investigated.

Finally, we wish to remark that there are still other factors that one might consider in choosing a scanning direction. For example,[2] in skew scanning, the lines are not of equal length, therefore the power of the video signal does not have peaks at multiples of line frequency. Hence, when several video signals share the same channel, the use of skew scanning will reduce cross modulation. On the other hand, skew scanning complicates line synchronization.

T. S. Huang

References

1. T. S. Huang, Two-dimensional power density spectrum of television random noise, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M. I. T., April 15, 1963, p. 147.

2. Dr. B. Prasada, Bell Telephone Laboratories, Inc., Private communication, 1964.

2. BOUNDS ON TWO-ELEMENT-KIND IMPEDANCE FUNCTIONS

In a previous report,[1] we discussed some bounds on the impedance functions of R, $\pm$L, $\pm$C, T networks. In this report, we shall present bounds for various types of two-element-kind impedance functions. We first prove a theorem for R, $\pm$C and R, $\pm$L networks.

THEOREM 1. Let $[Z_{ik}(s)]$ be an nth-order R,$\pm$C (or R,$\pm$L) open-circuit impedance matrix. Then $Z_{ik}(jw)$ satisfies

$$\left| Z_{ik}(jw) - \frac{1}{2}(Z_{ik}(0) + Z_{ik}(\infty)) \right| \leqslant \sqrt{\frac{1}{2}(Z_{ii}(0) - Z_{ii}(\infty)) \frac{1}{2}(Z_{kk}(0) - Z_{kk}(\infty))} \tag{1}$$

for all real w.

The proof of Theorem 1 follows readily from the two following lemmas.

LEMMA 1. Let $[Z_{ik}(s)]$ be an nth-order R,$\pm$C open-circuit impedance matrix. Then

$$Z_{ik}(s) = h_{ik}^{(\infty)} + h_{ik}^{(o)}/s^2 + \sum_{m=1}^{n} h_{ik}^{(m)}/(s+a_m), \tag{2}$$

where the real numbers a_m are independent of i and k, the $\left[h_{ik}^{(r)}\right]$ $(r=1, 2, \ldots, m; 0, \infty)$ are real and symmetrical, and

$\left[h_{ik}^{(\infty)}\right]$ is positive semidefinite (psd)

$\left[h_{ik}^{(o)}\right]$ is negative semidefinite (nsd)

$\left[h_{ik}^{(m)}\right]$ is psd if $a_m \geq 0$, and nsd if $a_m < 0$.

LEMMA 2. If $[p_{ik}]$ and $[q_{ik}]$ $(i, k=1, 2)$ are real, symmetrical, and psd, then

$$2p_{12}q_{12} \leq p_{11}q_{22} + q_{11}p_{22}. \tag{3}$$

By making appropriate impedance transformations, we deduce from Theorem 1 two theorems about $\pm R, C$ and $\pm R, L$ networks.

THEOREM 2. Let $[Z_{ik}(s)]$ be an nth-order $\pm R, C$ open-circuit impedance matrix. Then $Z_{ik}(jw)$ satisfies

$$\left| Z_{ik}(jw) - \tfrac{1}{2}(Z'_{ik}(0)/jw + Z'_{ik}(\infty)/jw) \right| \leq \sqrt{\tfrac{1}{2}(Z'_{ii}(0)/w - Z'_{ii}(\infty)/w)\tfrac{1}{2}(Z'_{kk}(0)/w - Z'_{kk}(\infty)/w)} \tag{4}$$

for any real w, where $Z'_{ik}(s) = sZ_{ik}(s)$.

THEOREM 3. Let $[Z_{ik}(s)]$ be an nth-order $\pm R, L$ open-circuit impedance matrix. Then $Z_{ik}(jw)$ satisfies

$$\left| Z_{ik}(jw) - \tfrac{1}{2}(jwZ''_{ik}(0) + jwZ''_{ik}(\infty)) \right| \leq \sqrt{\tfrac{1}{2}(wZ''_{ii}(0) - wZ''_{ii}(\infty))\tfrac{1}{2}(wZ''_{kk}(0) - wZ''_{kk}(\infty))} \tag{5}$$

for any real w, where $Z''_{ik}(s) = Z_{ik}(s)/s$.

Notice that inequalities (1), (4), and (5) are properties of the impedance functions and are independent of the manner in which one realizes these functions. When $i \neq k$, the inequalities give bounds on transfer functions; when $i = k$, they give bounds on driving-point functions.

It is clear that for RC(RL) networks, both Theorem 1 and Theorem 2 (Theorem 3) apply. For any particular realization, N, of an RC n-port, the quantities in (1) and (4) have the following physical interpretations:

$Z_{ik}(0)$ = open-circuit impedance matrix of N, when all capacitances are open-circuited

$Z_{ik}(\infty)$ = open-circuit impedance matrix of N, when all capacitances are short-circuited

$Z'_{ik}(0)/jw$ = open-circuit impedance matrix of N, when all resistances are short-circuited and $s = jw$

$Z'_{ik}(\infty)/jw$ = open-circuit impedance matrix of N, when all resistances are open-circuited and s = jw.

(For any RL n-port realization, we have similar physical interpretations.) The quantities $Z_{ii}(0)$, $Z_{ii}(\infty)$, $Z'_{ii}(0)$, and $Z'_{ii}(\infty)$ are not independent. In fact, we have the following lemmas.

LEMMA 3. $Z_{ii}(0)$ is finite, if and only if $Z'_{ii}(0)$ is zero.

LEMMA 4. $Z_{ii}(\infty)$ is zero, if and only if $Z'_{ii}(\infty)$ is finite.

In order to get useful bounds, one would like the right-hand sides of (1) and (4) to be finite. Hence, one would like to have $Z_{ii}(\infty) = 0 = Z'_{ii}(0)$. One can achieve this by the following procedure. Given an RC open-circuit impedance matrix $[Z_{ik}(s)]$, we form a new RC open-circuit impedance matrix

$$[z_{ik}(s)] = [Z_{ik}(s)] - [Z_{ik}(\infty)] - [Z'_{ik}(0)]/s. \qquad (6)$$

Then $z_{ii}(\infty) = 0 = z'_{ii}(0)$, where $z'_{ii}(s) = sz_{ii}(s)$, and we can apply inequalities (1) and (4) to $z_{ik}(s)$. For any nonzero finite w, the right-hand sides of (1) and (4) are finite, and $z_{ik}(jw)$ must lie in the intersection of two nondegenerate closed circular disks.
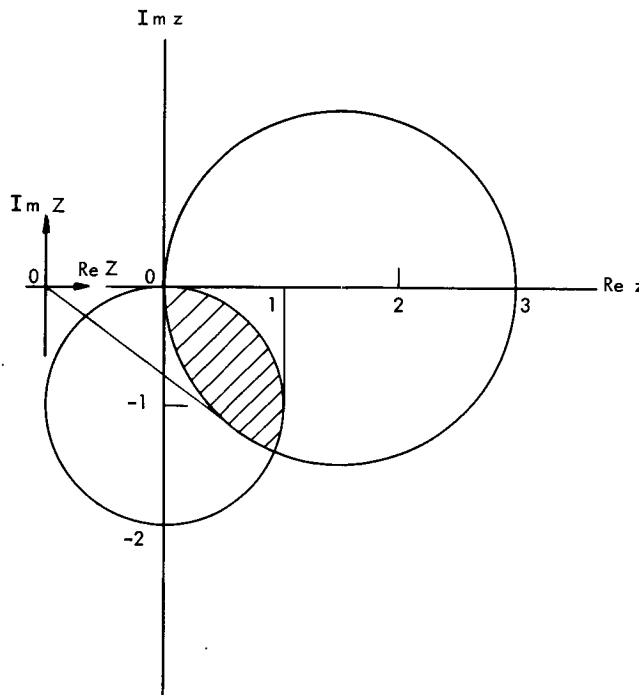
Fig. XVII-7.   Example illustrating Theorems 1 and 2.

We conclude with an example. Consider the RC driving-point impedance function $Z(s) = (s+2)(s+6)/(s+1)(s+3)$. Then $Z(0) = 4$ and $Z(\infty) = 1$, and inequality (1) implies $|Z(jw) - 5/2| \leqslant 3/2$. Let $Z'(s) = sZ(s)$. Then $Z'(0) = 0$ and $Z'(\infty) = \infty$. Therefore, inequality (4), when applied to $Z(s)$ directly, does not give any useful bounds. We can, however, define $z(s) = Z(s) - Z(\infty) - Z'(0)/s = 4(s+9/4)/(s+1)(s+3)$. Then $z(0) = 3$, $z(\infty) = 0$, $z'(0) = 0$, and $z'(\infty) = 4$. Hence for $w = 2$, say, $Z(j2)$ must lie in the shaded region of Fig. XVII-7. In particular, we have

$$1 \leqslant |Z(j2)| < 2.4, \qquad -37° < [\angle Z(j2)] \leqslant 0;$$

$$1 < \text{Re } Z(j2) < 2.1, \qquad -1.4 < \text{Im } Z(j2) \leqslant 0.$$

Putting $s = j2$ in the exact expression for $Z(s)$, we find $Z(j2) = 2.22 \underline{/-33.8°}$.

In the previous report,[1] we proved that if $[Z_{ik}(s)]$ is the open-circuit impedance matrix of an $R, \pm L, \pm C, T$ network, then

$$\left| Z_{ik}(jw) - \frac{R_{iko} + R_{iks}}{2} \right| \leqslant \sqrt{\left(\frac{R_{iio} - R_{iis}}{2} \right) \left(\frac{R_{kko} - R_{kks}}{2} \right)}, \qquad (7)$$

where $[R_{iko}]$ is the open-circuit impedance matrix of the network when all reactive elements are open-circuited, and $[R_{iks}]$ is the open-circuit impedance matrix of the network
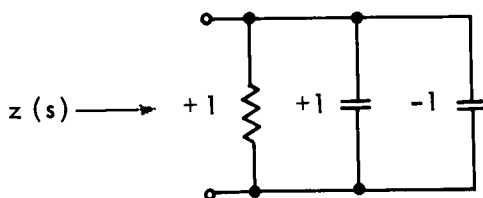
Fig. XVII-8. Example of an $R, \pm C$ network.

when all reactive elements are short-circuited.

We remark that Eq. 1 does not follow from Eq. 7, since, in general, for an $R, \pm C$ network, $Z_{ik}(0) \neq R_{iko}$ and $Z_{ik}(\infty) \neq R_{iks}$. For example, consider the network of Fig. XVII-8. We have $Z(s) = 1$; therefore, $Z(0) = 1 = Z(\infty)$. But $R_0 = 1$ and $R_s = 0$.

T. S. Huang

References

1. T. S. Huang and H. B. Lee, Bounds on impedance functions of $R, \pm L, \pm C, T$ networks, Quarterly Progress Report No. 75, Research Laboratory of Electronics, M. I. T., October 15, 1964, pp. 214-225.

C. SENSORY AIDS

1. APPROXIMATE FORMULAS FOR THE INFORMATION TRANSMITTED BY A DISCRETE COMMUNICATION CHANNEL

It is often desirable to have an approximate formula for the information transmitted by a discrete communication channel which is simpler than the exact expression.[1] In this report, two approximate expressions are derived. The derivations are instructive, for they show why two systems that operate with the same probability of error can have quite different information transmission capabilities.

Preliminary Theorems

The following theorems will be required. The proofs of Theorem 1 and of the lemma are omitted. Theorem 2 follows directly from Theorem 1, and also from Fano's discussion.[2]

THEOREM 1: Let $x_1, x_2, \ldots, x_n$ be non-negative real numbers. If $F(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_i \log x_i$, and if

$$\sum_{i=1}^{n} x_i = p,$$

then

$$F(p/n, p/n, \ldots, p/n) \leqslant F(x_1, x_2, \ldots, x_n) \leqslant F(p, 0, 0, \ldots 0).$$

The equality sign on the left applies only if all the x's are equal. The equality sign on the right applies only if all but one of the x's are zero.

THEOREM 2: Define

$p(x_i)$, probability of occurrence of the input x_i to a communication channel,

$p(y_j)$, probability of occurrence of the output y_j from a channel,

Lx, number of inputs having nonzero probability of occurrence, and

Ly, number of outputs having nonzero probability of occurrence.

Let

$$P(e \mid y_j) = 1 - p(x_j \mid y_j).$$

Then

$$P(e) = \sum_{j=1}^{Ly} [1 - p(x_j | y_j)] \, p(y_j) = \sum_{j=1}^{Ly} P(e | y_j) \, p(y_j).$$

If

$$H(e | y_j) = - [P(e | y_j) \log P(e | y_j) + [1 - P(e | y_j)] \log [1 - P(e | y_j)]],$$

$$H(e | Y) = \sum_{j=1}^{Ly} H(e | y_j) \, p(y_j),$$

and

$$H(e) = - P(e) \log P(e) - (1 - P(e)) \log (1 - P(e)),$$

then

$$0 \leq H(e | Y) \leq H(e).$$

LEMMA: Let $[p(y | x)]$ be a conditional probability matrix having Lx rows and Ly columns. Consider the matrix $[p(x | y)]$, where

$$p(x_i | y_j) = \frac{p(y_j | x_i) \, p(x_i)}{p(y_j)}.$$

If Q_j denotes the number of nonzero off-diagonal terms in the jth column of the matrix $[p(y | x)]$, then the number of nonzero off-diagonal terms in the jth row of the matrix $[p(x | y)]$ is also Q_j.

Derivation of Upper and Lower Bounds for I(X; Y)

To derive the following bounds on $I(X; Y)$ two different communication channels are considered, each of which is required to transmit information about the same input ensemble. Both channels have the same number of outputs, The two channel matrices have identical elements on the main diagonal. Therefore, $P(e | y_j)$ $(j = 1, 2, \ldots, Ly)$ and $P(e)$ are the same for both channels.

One channel matrix has only one nonzero off-diagonal term in each column. The information transmitted by this channel is a maximum for fixed values of $P(e | y_j)$ $(j = 1, 2, \ldots, Ly)$ and is equal to the upper bound of Eq. 2.

The other channel has a matrix in which all nonzero off-diagonal terms in any one column are equal. The information transmitted by this channel is a minimum for a given number of nonzero off-diagonal terms in each column, and for fixed values of $(Pe | y_j)$ $(j = 1, 2, \ldots, Ly)$. The information transmitted in this case is equal to the lower bound in Eq. 3.

THEOREM 3: Let $I(X; Y)$ be the information transmitted by a discrete channel, and

$$H(X) = - \sum_{i=1}^{Lx} p(x_i) \log p(x_i).$$

Let Q_{max} be the largest of Q_1, Q_2, ..., Q_{Ly}. Then

$$I(X; Y) \leq H(X) - H(e|Y) \leq H(X);$$

$$I(X; Y) \geq H(X) - H(e|Y) - \sum_{j=1}^{Ly} P(e|y_j) p(y_j) \log Q_j \geq H(X) - H(e) - P(e) \log Q_{max}.$$

PROOF:

$$I(X; Y) = \sum_{i=1}^{Lx} \sum_{j=1}^{Ly} p(x_i|y_j) p(y_j) \log \frac{p(x_i|y_j)}{p(x_i)}$$

$$= H(X) + \sum_{j=1}^{Ly} p(y_j) \sum_{i=1}^{Lx} p(x_i|y_j) \log p(x_i|y_j) \tag{1}$$

$$\sum_{j=1}^{Ly} p(y_j) \sum_{i=1}^{Lx} p(x_i|y_j) \log p(x_i|y_j) = \sum_{j=1}^{Ly} p(y_j) \left[p(x_j|y_j) \log p(x_j|y_j) + \sum_{\substack{i=1 \\ i \neq j}}^{Lx} p(x_i|y_j) \log p(x_i|y_j) \right].$$

If we replace x_i by $p(x_i|y_j)$, p by $\sum_{i=1}^{Lx} p(x_i|y_j)$, and n by Q_j, and if we use

$$1 - p(x_j|y_j) = \sum_{\substack{i=1 \\ i \neq j}}^{Lx} p(x_i|y_j),$$

then the inequalities

$$[1 - p(x_j|y_j)] \log \frac{[1 - p(x_j|y_j)]}{Q_j} \leq \sum_{\substack{i=1 \\ i \neq j}}^{Lx} p(x_i|y_j) \log p(x_i|y_j) \leq [1 - p(x_j|y_j)] \log [1 - p(x_j|y_j)]$$

follow directly from Theorem 1 and the lemma.

The equality sign on the right applies if, and only if, there is only one nonzero off-diagonal term in the jth row of the $p(x|y)$ matrix. The equality sign on the left applies

if, and only if, all the nonzero off-diagonal terms in the jth row of the $[p(x|y)]$ matrix are equal.

Substitution of the inequalities above in Eq. 1 results in

$$I(X; Y) \leq H(X) + \sum_{j=1}^{Ly} p(y_j) [p(x_j|y_j) \log p(x_j|y_j) + [1 - p(x_j|y_j)] \log [1 - p(x_j|y_j)]]$$

$$\leq H(X) - H(e|Y). \tag{2}$$

$$I(X; Y) \geq H(X) + \sum_{j=1}^{Ly} p(y_j) \left[p(x_j|y_j) \log p(x_j|y_j) + [1 - p(x_j|y_j)] \log \frac{[1 - p(x_j|y_j)]}{Q_j} \right]$$

$$\geq H(X) - H(e|Y) - \sum_{j=1}^{Ly} P(e|y_j) p(y_j) \log Q_j. \tag{3}$$

Theorem 3 now follows from Theorem 2 and from the fact that

$$\log Q_{max} \geq \log Q_j \qquad (j = 1, 2, \ldots, Ly).$$

Approximate Formula for I(X; Y)

In order to use upper and lower bounds to estimate $I(X; Y)$ in such a way that the expected value of the estimation error is minimized, it is necessary to know the distribution function for $I(X; Y)$. Since the distribution function is not usually available, the estimate for $I(X; Y)$ will be taken as the average of its upper and lower bounds. Such an estimate minimizes the maximum possible estimation error.

It follows that we estimate that

$$I_1(X; Y) = H(X) - H(e|Y) - \frac{1}{2} \sum_{j=1}^{Ly} P(e|y_j) p(y_j) \log Q_j \tag{4}$$

$$I_2(X; Y) = H(X) - \frac{1}{2} (H(e) + P(e) \log Q_{max}). \tag{5}$$

The maximum estimation error e is given in each case by

$$e_1 = \frac{1}{2} \sum_{j=1}^{Ly} P(e|y_j) p(y_j) \log Q_j \tag{6}$$

$$e_2 = \frac{1}{2} (H(e) + P(e) \log Q_{max}). \tag{7}$$

The maximum error e in per cent, which results because the estimate for I was chosen midway between the upper and lower bound, is

$$e\% = \left(\frac{U-L}{2L}\right) 100\%,$$

where U is the upper bound, and L is the lower bound. Thus

$$e_1\% = \left[\frac{\frac{1}{2}\sum_{j=1}^{Ly} P(e\,|\,y_j)\,p(y_j)\,\log Q_j}{H(X) - H(e\,|\,Y) - \sum_{j=1}^{Ly} P(e\,|\,y_j)\,p(y_j)\,\log Q_j}\right] 100\%, \tag{8}$$

$$e_2\% = \frac{\frac{1}{2}(H(e) + P(e)\,\log Q_{max})}{H(X) - H(e) - P(e)\,\log Q_{max}}\,100\%. \tag{9}$$

The use of inequalities

$$H(e\,|\,Y) \leqslant H(e)$$

and

$$\sum_{j=1}^{Ly} P(e\,|\,y_j)\,p(y_j)\,\log Q_j \leqslant P(e)\,\log Q_{max}$$

in (6) and (8) results in upper bounds for e_1 and $e_1\%$

$$e_1 \leqslant \frac{1}{2}P(e)\,\log Q_{max} \tag{10}$$

$$e_1\% \leqslant \left(\frac{\frac{1}{2}P(e)\,\log Q_{max}}{H(X) - H(e) - P(e)\,\log Q_{max}}\right) 100\%, \tag{11}$$

which are easier to compute than the exact quantities given by Eqs. 6 and 8.

In Figs. XVII-9 and XVII-10 $e_1\%$ and $e_2\%$ are plotted as functions of P(e) for various values of Q_{max} for the cases H(X) = 4 and H(X) = 7. It should be remembered that these graphs represent the maximum errors that can occur as a result of approximating I(X; Y) by $I_1(X; Y)$ and $I_2(X; Y)$. The actual error that results when I(X; Y) is approximated by $I_2(X; Y)$ will equal the maximum error if and only if
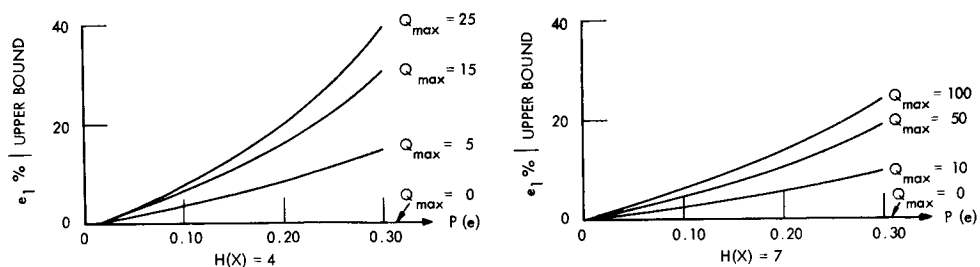
Fig. XVII-9. $e_1\%$ | upper bound vs P(e) and Q_{max}.
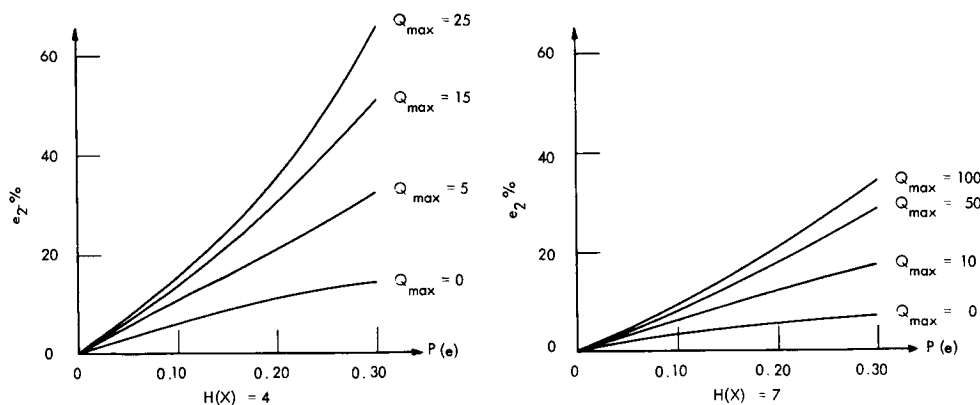
Fig. XVII-10. $e_2\%$ vs P(e) and Q_{max}.

$$I(X; Y) = H(X), \tag{12}$$

or

$$I(X; Y) = H(X) - H(e) - P(e) \log Q_{max}. \tag{13}$$

Equation (12) holds if, and only if, all off-diagonal terms in the channel matrix are zero (a perfect communication system). Equation (13) applies if, and only if, $H(e|y_j) = H(e)$, and $Q_j = Q_{max}$ $(j = 1, 2, \ldots, Ly)$(the same number of off-diagonal terms in each column of the channel matrix, and all these terms equal).

Similarly, the errors that result when I(X; Y) is approximated by $I_1(X; Y)$ are equal to the maximum error only in special cases. If there is only one off-diagonal term in each column of the channel matrix, then

$$I(X; Y) = H(X) - H(e|Y).$$

If all the off-diagonal terms in each column are equal, then

$$I(X;Y) = H(X) - H(e \mid Y) - \sum_{j=1}^{Ly} P(e \mid y_j) \, p(y_j) \, \log Q_j.$$

The estimate $I_1(X;Y)$ is always better than or as good as $I_2(X;Y)$. However, the first estimate requires more computation than the latter. A useful procedure for estimating $I(X;Y)$ is:

1. Evaluate e_2 in per cent. If e_2 is acceptable, evaluate $I_2(X;Y)$ as an approximation to $I(X;Y)$.

2. If e_2 is too large, evaluate the upper bound for e_1 in per cent. If this upper bound is acceptable, evaluate $I_1(X;Y)$ as an approximation to $I(X;Y)$.

3. If the upper bound to e_1 is too large, then compute $I(X;Y)$ from the exact formula (1).

Example

The following channel matrix results when a human subject is required to make one of eight responses to one of eight equiprobable statistically independent stimuli. The information transmitted is to be computed to an accuracy of ± 5 per cent of the true value.

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	
x_1	.95	.05							
x_2	.05	.90	.05						
x_3	.05	.05	.90						
x_4			.10	.80	.10				$= [P(y \mid x)].$
x_5					.90	.05		.05	
x_6						.95	.05		
x_7						.05	.90	.05	
x_8						.05		.05	.90	

Step 1: Computation of e_2 (per cent)

$Q_{max} = 2$

$P(e) = 0.10$

$H(X) = 3$

$e_2 = 11.9\%.$

The maximum error (per cent) resulting from the simpler estimate exceeds the desired ±5 per cent bound.

Step 2: Computation of e_1 (per cent)

$$e_1 \leqslant 2.2\%.$$

The maximum percentage in error that is caused by using $I_1(X;Y)$ as an estimate for $I(X;Y)$ is within the required limits of accuracy.

$$\frac{1}{2} \sum_{j=1}^{8} P(e|y_j) \, p(y_j) \, \log Q_j = 0.050$$

$$H(e|Y) = 0.383$$

$$I_1(X;Y) = 2.57 \text{ bits/stimulus.}$$

An exact calculation shows that

$$I(X;Y) = 2.59 \text{ bits/stimulus.}$$

Discussion

The amount of computation required for the estimate $I_1(X;Y)$ increases in proportion to the number of messages. The simpler estimate requires little computation and is independent of the number of messages. The maximum error (per cent) for both estimates decreases as $H(X)$ increases, since the influence of $H(e)$ in the denominator of equations (8) and (9) becomes less as $H(X)$ becomes larger. When $H(X)$ is small, the first estimate will usually be required. For larger values of $H(X)$, the second estimate will usually yield acceptable values of e_2 per cent. While it is true that the amount of computation necessary for the evaluation of $I_1(X;Y)$ increases with the number of messages, it is also true that the probability that the simpler estimate will be satisfactory also increases with $H(X)$.

R. W. Donaldson

References

1. R. M. Fano, Transmission of Information (The M. I. T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961).

2. Ibid., pp. 46-47.

Prof. M. Eden
Prof. W. T. Peake[†]
Prof. R. R. Pfeiffer[†]
Prof. W. A. Rosenblith
Prof. W. M. Siebert
Prof. T. F. Weiss[†]
Dr. J. S. Barlow[‡]
Dr. A. W. B. Cunningham[**]
Dr. E. Gibermann[††]
Dr. R. D. Hall
Dr. N. Y-s. Kiang[†]
Dr. R. R. Rojas Corona
Dr. R. Shofer[‡‡]
Dr. W. Simon[***]
Dr. E. B. Vidale
J. A. Anderson
JoAnn Arnn

L. Braida
R. M. Brown[†]
S. K. Burns
R. J. Clayton
A. H. Crist[†]
N. I. Durlach
J. A. Freeman
P. R. Gray
J. J. Guinan, Jr.[†]
F. N. Jordan
L. J. Krakauer
D. P. Langbein[†]
P. H. Levine
P. L. Marcus
R. G. Mark
E. G. Merrill

S. A. Miller
D. C. Milne
C. E. Molnar
M. J. Murray
M. Nahvi
P. H. O'Lague
Ann M. O'Rourke
L. Peusner
D. J. M. Poussart
J-C. Proteau
L. R. Rabiner
M. B. Sachs
M. M. Scholl
J J. Singer
Judith E. Swensen
I. H. Thomae
M. L. Wiederhold

A. SUMMARY OF RESEARCH — PUBLICATIONS

The following list of publications and theses presented by members of the Communications Biophysics Group during 1964 was omitted from Quarterly Progress Report No. 76, January 15, 1965.

Publications

1. J. S. Barlow, Evoked Responses in Relation to Visual Perception and Oculomotor Reaction Times in Man, Ann. N.Y. Acad. Sci. 112, 432-467 (1964).

2. J. S. Barlow, Some Statistical Characteristics of Electrocortical Activity in Relation to Visual-Oculomotor Tracking in Man, First Conference on Neurobiology: Feedback Systems Controlling Nervous Activities, Alfonso Escobar (ed.), Asociedad Mexicana de Ciencias Fisiologicas: Mexico City, D.F., 1964, pp. 385-408; Boletín del Instituto de Estudios Médicos y Biológicos, 21, 497-517 (1963).

3. J. S. Barlow, R. L. Rovit, and P. Gloor, Correlation Analysis of EEG Changes Induced by Unilateral Intracarotid Injection of Amobarbital, Electroenceph. clin. Neurophysiol. 16, 213-220 (1964).

Publications (continued)

4. M. A. B. Brazier, Evoked Responses Recorded from the Depths of the Human Brain, Ann. N.Y. Acad. Sci. 112, 33-59 (1964).

5. M. A. B. Brazier, Information Carrying Characteristics of Brain Response, The Physiological Basis of Mental Activity, R. Hernandez-Peon (ed.), Suppl. 24, Electroenceph. clin. Neurophysiol., 1963, pp. 55-67.

6. A. Cavaggioni and M. H. Goldstein, Jr., Facilitation and Inhibition in the Visual System after Photic Stimulation, Archives Italiennes de Biologie (in press).

7. N. I. Durlach, Note on Binaural Masking-Level Differences at High Frequencies, J. Acoust. Soc. Am. 36, 576-581 (1964).

8. N. I. Durlach, Note on Binaural Masking-Level Differences as a Function of the Interaural Correlation of the Masking Noise, J. Acoust. Soc. Am. 36, 1613-1617 (1964).

9. M. Eden, A Note on Short-Term Storage of Information in Vision, Perceptual and Motor Skills 19, 93-94 (1964).

10. M. Eden, Taxonomies of Disease, The Diagnostic Process, John A. Jacques (ed.), Proceedings of a conference sponsored by the Biomedical Data Processing Training Program, University of Michigan, Ann Arbor, Michigan, April 1964, pp. 47-49.

11. Denise Albe-Fessard, J. Masson, Robert Hall, et Walter Rosenblith, Modifications au cours de la veille et du sommeil des valeurs moyennes de réponses nerveuses centrales induites par des stimulations somatiques chez le Chat libre, Compt. Rend. Acad. Sci. (Paris) 258, 353-356 (6 janvier 1964).

12. G. L. Gerstein and W. A. Clark, Simultaneous Studies on Firing Patterns in Several Neurons, Science 143, 1325-1327 (1964) (Abstract).

13. G. L. Gerstein and N. Y-s. Kiang, Responses of Single Units in the Auditory Cortex, Exptl. Neurol. 10, 1-18 (1964).

14. G. L. Gerstein and B. Mandelbrot, Random Walk Models for the Spike Activity of a Single Neuron, Biophys. J. 4, 41-68 (1964).

15. J. L. Hall II, Minimum Detectable Change in Interaural Time or Intensity Difference for Brief Impulsive Stimuli, J. Acout. Soc. Am. 36, 2411-2413 (1964).

16. N. Y-s. Kiang, Stimulus Coding in the Auditory Nerve and Cochlear Nucleus, Acta Oto-Laryngologica (in press).

17. N. Y-s. Kiang and R. R. Pfeiffer, Nerve and Nucleus: A Study of Stimulus Coding in the Initial Stages of the Auditory Nervous System, Science 146, 432 (1964) (Abstract).

18. R. R. Pfeiffer, Response Characteristics of Some Single Units in the Cochlear Nucleus to Tone-Burst Stimulation, J. Acoust. Soc. Am. 36, 1017 (1964) (Abstract).

19. R. R. Pfeiffer and N. Y-s. Kiang, Patterns of Spontaneous and Continuously Stimulated Spike Discharges in the Cochlear Nucleus of Anesthetized Cats, Biophys. J. (in press).

20. M. B. Sachs, Responses to Acoustic Stimuli from Single Units in the Eight Nerve of the Green Frog, J. Acoust. Soc. Am. 36, 1956-1958 (1964).

21. W. M. Siebert, Some Implications of the Stochastic Behavior of Primary Auditory Neurons, Kybernetik (in press).

22. D. C. Teas and N. Y-s. Kiang, Evoked Responses from the Auditory Cortex, Exptl. Neurol. 10, 91-119 (1964).

Theses

1. Richard S. Bair, Pitch of Short Tones, S. B. Thesis, Department of Physics, M. I. T.

2. Gerald J. Burnett, Masking Effects of Pure Tones in the Human Ear, S. B. Thesis, Department of Electrical Engineering, M. I. T.

3. Stephen K. Burns, Neuroelectric correlates of Behavioral Abnormalities Induced by Early Sensory Restriction, S. M. Thesis, Department of Electrical Engineering, M. I. T.

4. Robert R. Capranica, The Evoked Vocal Response of the Bullfrog — A Study of Communication by Sound, Sc. D. Thesis, Department of Electrical Engineering, M. I. T. (to be published as Special Technical Report Number 12 of the Research Laboratory of Electronics and as Research Monograph No. 33 by The M. I. T. Press, Cambridge, Mass.).

5. Ambrose W. Clay, The Effects of Clutter on the Echolocation System of Bats, S. B. Thesis, Department of Electrical Engineering, M. I. T.

6. Richard J. Clayton, Rhythmic Forebrain Potential Correlated with Cessation of Movement in the Rat, S. M. Thesis, Department of Electrical Engineering, M. I. T.

7. Harry S. Colburn, Time-Intensity Relations for Binarual Masking, S. M. Thesis, Department of Electrical Engineering, M. I. T.

8. Edwin G. Duffin, Electrical Properties of Indium-Filled Microelectrodes, S. B. Thesis, Department of Electrical Engineering, M. I. T.

9. Daniel S. Frischmuth, Effects of Control of Transmitted Signal Reception on Human Echolocation, S. B. Thesis, Department of Electrical Engineering, M. I. T.

10. John J. Guinan, Jr., The Transfer Characteristic of the Cat's Middle Ear, S. M. Thesis, Department of Electrical Engineering, M. I. T.

11. Harold W. Ingels, Noise Masked Threshold of Periodic Pulses, S. M. Thesis, Department of Electrical Engineering, M. I. T.

12. Kenneth C. Koerber, Spontaneous Activity in the Cochlear Nuclei of De-cochleated Cats, S. M. Thesis, Department of Electrical Engineering, M. I. T.

13. Clifford L. Laurence, Studies of Masking Noise Phase Delays in Binaural Masking Level Differences, S. B. Thesis, Department of Physics, M. I. T.

14. James P. McGaughy, Auditory Localization of Signals in the Presence of Masking Noise, S. B. Thesis, Department of Electrical Engineering, M. I. T.

15. Donald J. Mided, The Role of Psychophysics in the Law, S. B. Thesis, Department of Electrical Engineering, M. I. T.

16. Michael J. Murray, Analysis and Synthesis of the Mating Croak of the Bullfrog, S. B. Thesis, Department of Electrical Engineering, M. I. T.

17. John T. Motor, Binaural Detection of Signal with Angular Dispersion of Masking Noise, S. B. Thesis, Department of Electrical Engineering, M. I. T.

18. Richard F. Otte, A Miniature Microscope Extension Using Fiber Optics, S. M. Thesis, Department of Electrical Engineering, M. I. T.

19. Lawrence R. Rabiner, Binaural Masking — The Effects of Interaural Delay of the Noise on the Detection of Tones, S. M. Thesis, Department of Electrical Engineering, M. I. T.

20. Murray B. Sachs, Characteristics of Primary Auditory Neurons in the Green Frog. S. M. Thesis, Department of Electrical Engineering, M. I. T.

Theses (continued)

21. Abba Weinstein, Measurement of the JND in Intensity, S. B. Thesis, Department of Electrical Engineering, M. I. T.

22. George E. Wien, A Preliminary Investigation of the Effect of Head Width on Binaural Hearing, S. M. Thesis, Department of Electrical Engineering, M. I. T.

W. M. Siebert, W. A. Rosenblith

B. MOTION OF MIDDLE-EAR JOINTS

To determine whether there is any significant articulation of the middle-ear joints, we measured the phase and amplitude of three points, (i) on the body of the stapes, (ii) on the long process of the incus, and (iii) near the end of the handle of the malleus. The measurements were made on anesthetized cats by a method described earlier.[1,2] The phase delays and amplitudes of the stapes and incus relative to the malleus are plotted in Fig. XVIII-1.

Most investigators (see, for example, Møller[3] and Zwislocki[4]) have stated that the malleus and incus move as a unit, whereas the stapes moves separately with slippage occurring in the incudo-stapedial joint. The data indicate, however, that there is movement in the malleo-incudal joint, and little, if any, in the incudo-stapedial joint. Figure XVIII-1 shows a slight increase in the incus
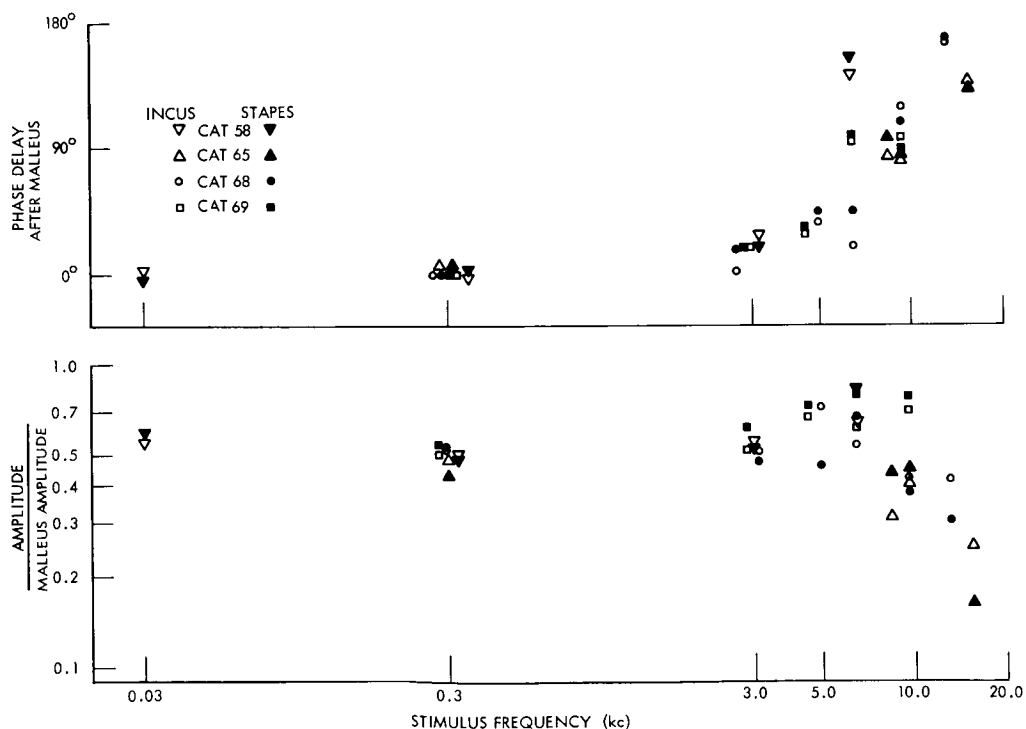
Fig. XVIII-1. Phase and amplitude of the incus and stapes relative to the malleus.

and stapes movement (relative to the malleus), in the 4-9 kc region, a decrease in the relative movement at higher frequencies, and an increase of the relative phase delay above 4 kc. All of these observations agree with a model of the middle ear in which the malleus and incus are coupled elastically.

J. J. Guinan, Jr., W. T. Peake

References

1. J. J. Guinan, Jr., The Transfer Characteristic of the Cat's Middle Ear, S. M. Thesis, Department of Electrical Engineering, M. I. T., 1964.

2. J. J. Guinan, Jr. and W. T. Peake, Motion of middle-ear bones, Quarterly Progress Report No. 74, Research Laboratory of Electronics, M. I. T., July 15, 1964, pp. 219-221.

3. A. Møller, Network model of the middle ear, J. Acoust. Soc. Am. 33, 168 (1961).

4. J. Zwislocki, Analysis of the middle-ear function, Part I: Input impedance, J. Acoust. Soc. Am. 34, 1514 (1962).

C. DEPENDENCE OF EFFERENT INHIBITION OF AUDITORY NERVE RESPONSES ON INTENSITY OF ACOUSTIC STIMULI

Continuation of earlier studies[1] on the efferent olivocochlear bundle (OCB) has confirmed the impression that the effect of electrical stimulation of the crossed OCB is dependent on the strength of the acoustic stimuli. In particular, the data presented here indicate that stimulation of the OCB, which produces considerable reduction in the amplitude of auditory nerve responses to low-intensity clicks, has little, if any, effect on the neural responses to high-intensity clicks. This finding differs from results published by Desmedt.[2]

The data presented here were obtained from anesthetized cats with both middle-ear muscles cut. Responses were recorded from an electrode near the round window (reference on headholder) and from bipolar concentric electrodes in the auditory nerve. In all of the results presented here electrical stimulation of the crossed OCB consisted of a burst of 32 shocks at a rate of 400 per second. The pulse that produced the click was delayed 2 msec from the last of the shocks.

Figure XVIII-2 shows averaged responses recorded from the round-window electrode over a range of 80 db. Figure XVIII-3 is a similar presentation of responses recorded from the concentric electrode in the auditory nerve.

Figure XVIII-4 shows a plot of the amplitude of the first neural component, N_1, of the round-window click response as a function of click intensity. Figure XVIII-5 is a similar intensity series for the responses recorded from the concentric electrodes. It can be seen from Figs. XVIII-2 through XVIII-5 that throughout the upper 40 db of the

CONDENSATION CLICKS ROUND WINDOW ELECTRODE

500 μv PEAK-TO-PEAK, 1000 CPS

(a) -80 db AMPLIFICATION×5

(b) -60 db AMPLIFICATION×5

(c) -40 db AMPLIFICATION×2

(d) -20 db AMPLIFICATION×1

(e) 0 db AMPLIFICATION×1/2
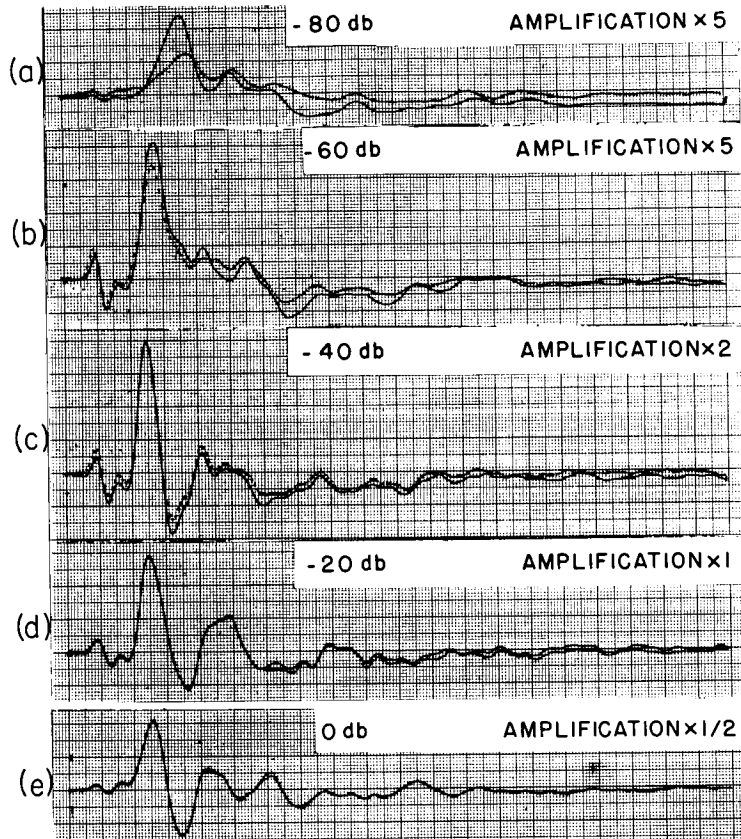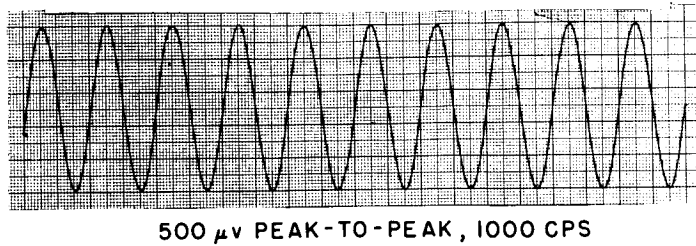
Fig. XVIII-2. Averaged responses to condensation clicks recorded at the round window. Solid-line traces are click responses with no OCB stimulation; dotted-line traces are responses to clicks presented 2 msec after the last of 32 shocks, 1.8 volts peak-to-peak, delivered at a rate of 400/sec to the crossed OCB at its decussation. Click intensity (re 4 volts to PDR-10 earphone, 0.1 msec square pulses) is indicated on each trace. The vertical scale amplification, relative to the calibration signal, accounts for number of responses averaged and changes in amplifier gain. Visual detection level for single responses, -90 db. Number of responses averaged (a) 64, (b), (c), (d), (e), 32.

CONDENSATION CLICKS AUDITORY NERVE ELECTRODE

500 μv PEAK-TO-PEAK, 1000 CPS

(a) -80 db   AMPLIFICATION × 10

(b) -60 db   AMPLIFICATION × 5

(c) -40 db   AMPLIFICATION × 2

(d) -20 db   AMPLIFICATION × 1

(e) 0 db   AMPLIFICATION × 1
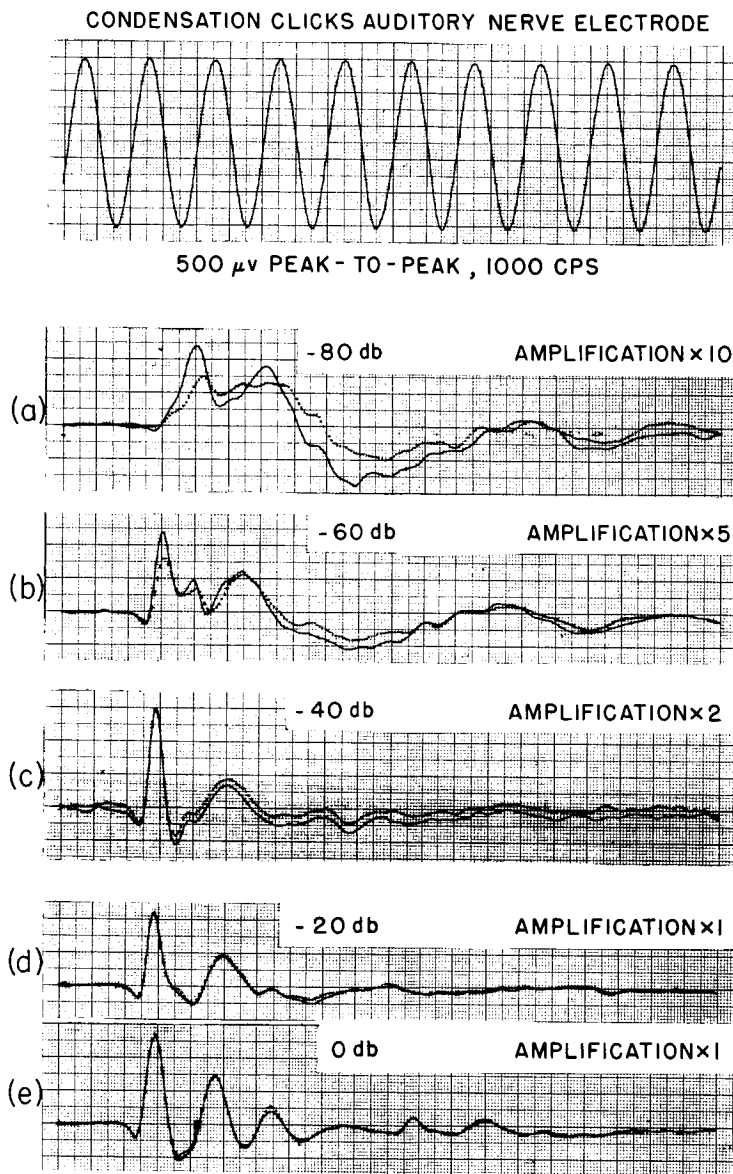
Fig. XVIII-3. Same as Fig. XVIII-2 except that responses (to same stimuli) were recorded from concentric bipolar electrode in the auditory nerve. Number of responses averaged (a) 64, (b), (c), (d), (e), 32.

Fig. XVIII-4.

Amplitude of N_1 component of round-window response to condensation clicks. Insert shows how N_1 was measured.

Data from same series shown in Fig. XVIII-2. Open circles represent response amplitude with no OCB stimulation; crosses, response amplitude with click preceded by OCB stimulation; open squares, N_1 response amplitude in a control series without OCB stimulation run immediately after the shock series at each intensity. Solid line represents the intensity series without efferent stimulation; dashed line, intensity series with OCB stimulation.

Fig. XVIII-5.

Same as Fig. XVIII-4 except that responses were recorded from concentric bipolar electrode in auditory nerve. Data from same series as shown in Fig. XVIII-3.

CONDENSATION CLICKS COCHLEAR MICROPHONIC RESPONSES

500 μv PEAK-TO-PEAK, 1000 CPS

-80 db
AMPLIFICATION × 40

(a)

-60 db
AMPLIFICATION × 10

(b)

-40 db
AMPLIFICATION × 4

(c)

-20 db
AMPLIFICATION × 4

(d)

0 db
AMPLIFICATION × 4

(e)
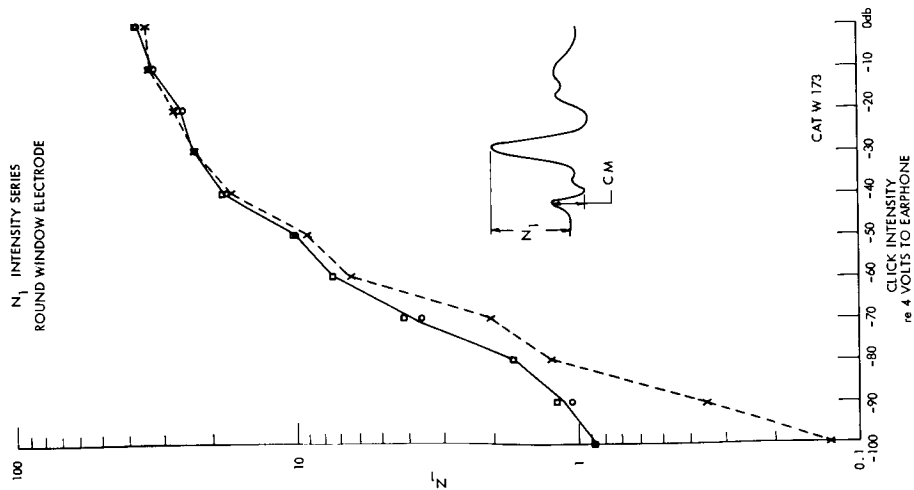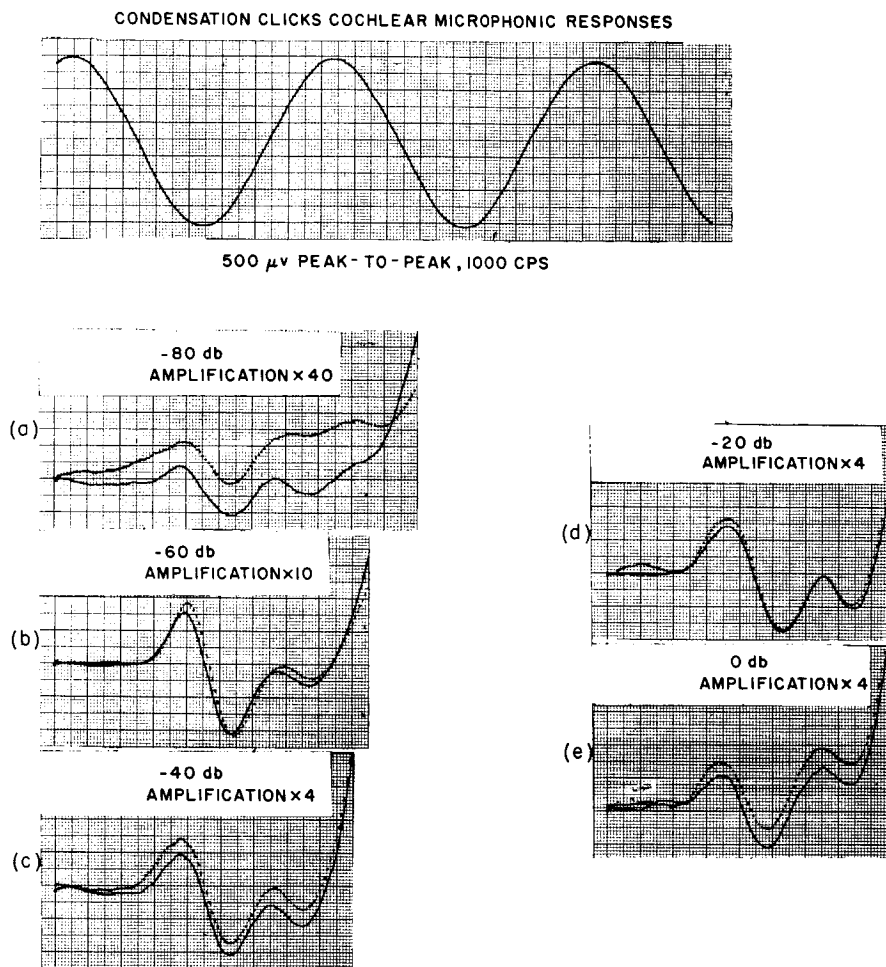
Fig. XVIII-6. Averages of same responses as shown in Fig. XVIII-2 with
time scale expanded by a factor of four and vertical scale
expanded as shown in each trace relative to calibration
signal. Solid line traces represent round-window response
with no OCB stimulation; dotted line traces, response to
click preceded by OCB stimulation. Number of responses
averaged, (a) 64, (b), (c), (d), (e), 32.

intensity range the OCB stimulation seems to have little effect on the neural part of the click response.

Figure XVIII-6 shows averaged responses from the round-window electrode with the scales expanded to better display the cochlear microphonic (CM) response. As has been reported[2,3] by other workers the CM component increases after OCB stimulation. Figure XVIII-7 is a plot of peak-to-peak amplitude of CM as a function of click intensity.
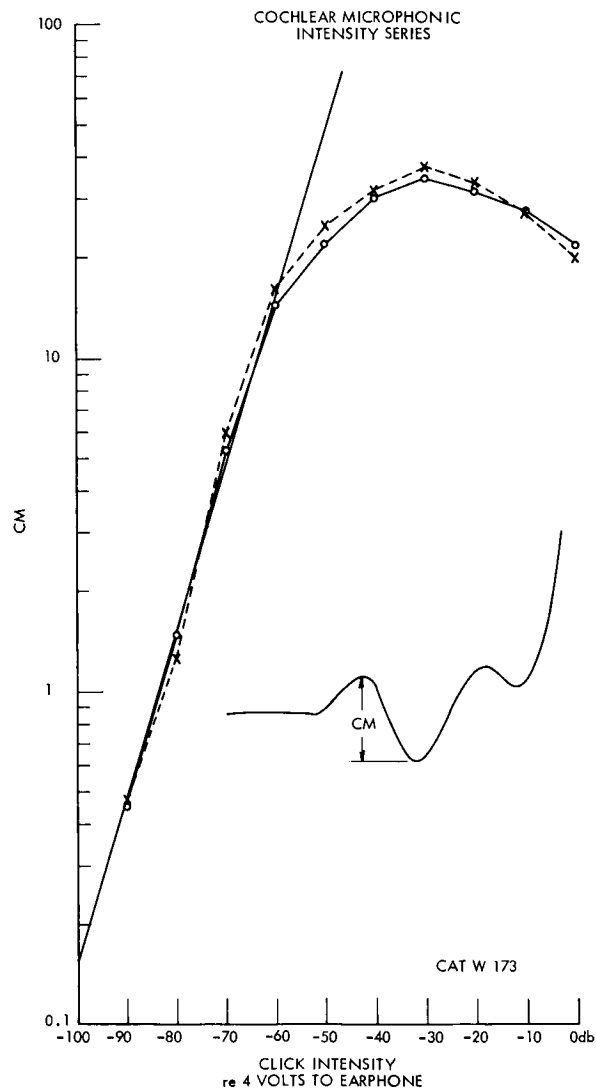
Fig. XVIII-7. Cochlear microphonic intensity series. Open circles, crosses, solid and dashed lines same as in Figs. XVIII-4 and XVIII-5. CM amplitude measured peak-to-peak as indicated in insert. (Fine line indicates a linear relationship between CM and sound-pressure level.)

The increase of CM from OCB stimulation can be seen to extend over a larger intensity range than does N_1 reduction. In other preparations this increase has existed throughout the intensity range.

Desmedt[2], in similar studies confined to low acoustic intensities, concluded that the effect of OCB stimulation on neural responses to clicks can be thought of as reducing the intensity of the acoustic stimulus. The amount of equivalent reduction, he concluded, is largely independent of the acoustic intensity for a given configuration of OCB stimulation. If this were true over a large intensity range, the dashed curves in Figs. XVIII-4 and XVIII-5 would be shifted to the right by a constant amount relative to the solid curves. The dashed and solid curves of Figs. XVIII-4 and XVIII-5 are clearly not parallel; there appears, rather, to be a continuous decrease of the OCB effect on N_1 with increasing acoustic intensity. At intensities 70 db or more above threshold for detection of N_1, the OCB seems to be incapable of reducing the neural response.

The effects demonstrated here can be enhanced somewhat by increasing the voltage of the electrical stimulus to the OCB. (In all experiments discussed here the shocks were at a level that caused no observable muscular response; no paralyzing drug was given.) It was observed, however, that even with shocks strong enough to cause gross movements of the preparation there was no evidence of OCB reduction of the N_1 response to 0 db clicks.

We have not yet been able to interpret these data in terms of possible characteristics of the mechanisms of action of the OCB. It appears from preliminary data that the frequency spectrum of the acoustic stimulus is also an important parameter in determining the effectiveness of efferent inhibition. The combination of frequency and intensity dependence is under investigation at the present time.

<div align="right">M. L. Wiederhold, W. T. Peake</div>

References

1. M. L. Wiederhold and Eleanor K. Chance, Effects of olivocochlear bundle stimulation on acoustically evoked potentials, Quarterly Progress Report No. 70, Research Laboratory of Electronics, July 15, 1963, p. 311.

2. J. E. Desmedt, Auditory-evoked potentials from cochlea to cortex as influenced by activation of the efferent olivocochlear bundle, J. Acoust. Soc. Am. 34, 1478 (1962).

3. J. Fex, Auditory activity in centrifugal and centripetal cochlear fibers in cat, Acta. Physiol. Scand. (Stockholm) 55, Suppl. 189 (1962).

D. CORRELATION ANALYSIS OF EEG AND TREMOR ACTIVITY

In the account of studies on normal or "physiological" tremors, which was included in the report[1] on investigation of tremors in Quarterly Progress Report No. 76, it was suggested that it would be of interest to compare by correlation techniques simultaneously recorded EEG and tremograms, in view of the appearance in the latter of components of the same frequencies as that of alpha activity (8-13/sec in the parieto-occipital EEG). In the present report findings are presented from analyses of such recordings from two normal controls (the same subjects as were studied in the previous report[1]) and in two patients, the first having a pronounced unilateral rhythmic tremor of the outstretched upper extremity, and the second having a rather rhythmic physiological tremor. The techniques of recording analysis of the data were the same as those described in the earlier report.[1] All recordings were made with eyes closed and were of 2 minutes duration, except for those carried out during intermittent photic stimulation, which recordings were of 40 sec duration.

In Fig. XVIII-8 are shown a portion of the ink trace and correlograms for the first normal control. The location of the first peak in the autocorrelograms of the tremograms at a delay of 100 msec is indicative of tremor activity of an average frequency of 10/sec. The last is very close to the average EEG frequency of 9.5/sec which is evidenced from the autocorrelogram for the latter. From the crosscorrelograms in Fig. XVIII-8, it is apparent that there is only minimal crosscorrelation between the tremors on the two sides. It is also apparent that the crosscorrelograms for the midline parieto-occipital EEG with the tremors on the two sides, are essentially flat.

Results for similar recordings from the second normal control subject, from whom bilateral EEG recordings in the parieto-occipital region were made, are shown in Fig. XVIII-9. From the autocorrelograms, it is apparent that the average tremor frequency, 8/sec, is much lower than that of the average EEG frequency, which is 11.1/sec. The crosscorrelation coefficient between the tremors from the two sides is seen to be 0.25. The crosscorrelation coefficient for the EEG on the two sides is approximately 0.6 at zero delay. The crosscorrelograms for the tremors with the EEG on the same and on opposite sides are essentially flat.

In Fig. XVIII-10, results are shown from recordings, carried out at the suggestion of Dr. G. F. Rossi of Genoa, Italy, during a visit to the Neurophysiology Laboratory of the Massachusetts General Hospital, to investigate the possibility of altering the frequency characteristics of the tremor by repetitive photic stimulation of the subject. It is apparent from the autocorrelograms in Fig. XVIII-10 that the stroboscopic stimulation was without effect upon tremor activity, in contrast to its effect upon the EEG. In all instances, crosscorrelograms of EEG with tremor activity were essentially flat, and hence none is shown in Fig. XVIII-10.
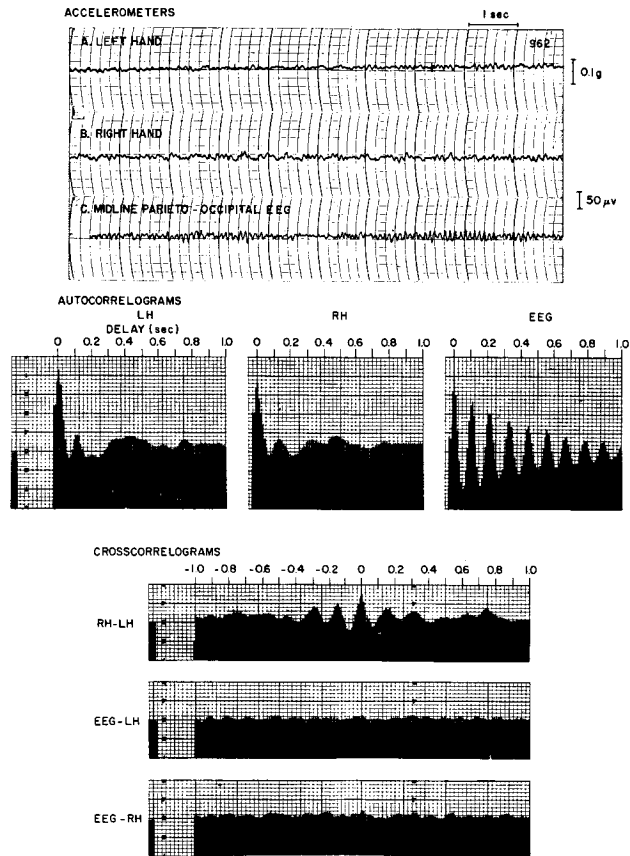
Fig. XVIII-8.   Simultaneously recorded midline parieto-occipital
EEG and accelerometrically monitored physio-
logical tremors of the outstretched upper extrem-
ities for a normal subject (B. J.). Accelerometers
mounted on the upper surface of the hands. Eyes
closed in this and subsequent figures. Duration
of this and subsequent recordings, 2 minutes,
except for recordings during intermittent photic
stimulation, for which 40-sec recordings were
made. Calibration is in terms of g, the acceler-
ation caused by gravity.
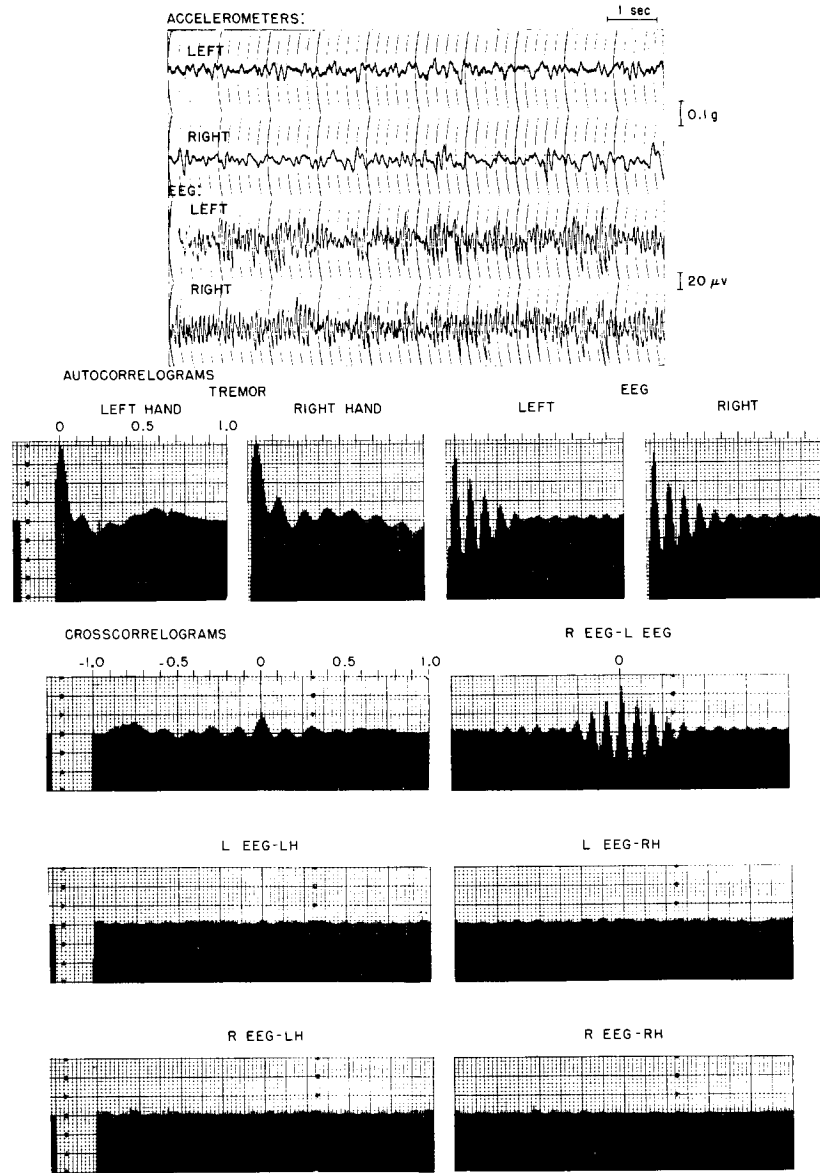
Fig. XVIII-9. Bilateral tremograms and parieto-occipital EEG
recordings for a normal subject (B. K.).
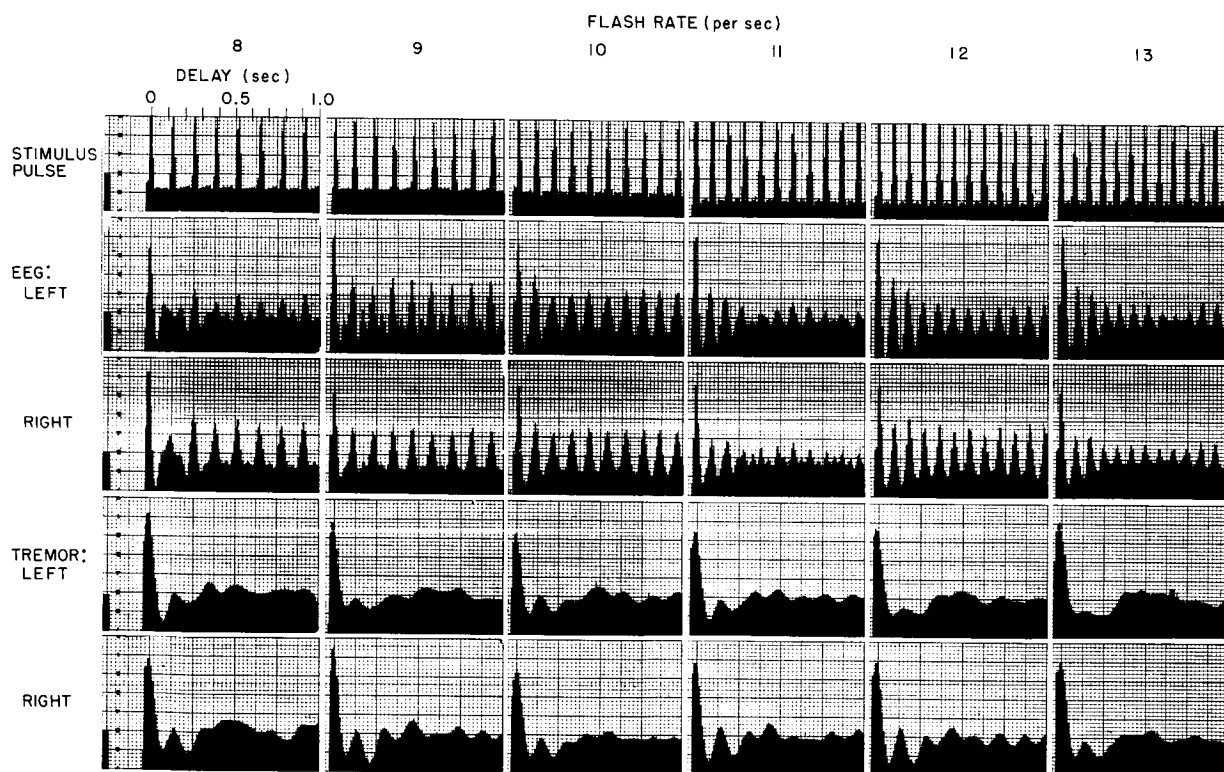
Fig. XVIII-10. Autocorrelograms for EEG and tremograms for intermittent photic
stimulation at various flash rates. (Normal Subject B. K.)
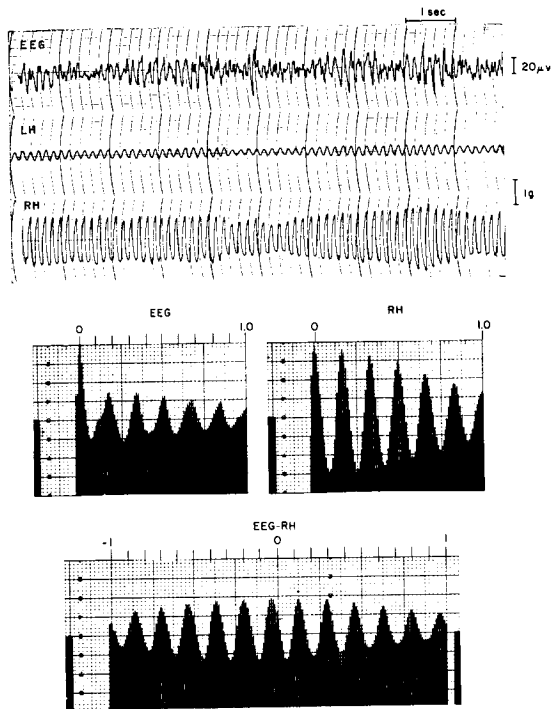
Fig. XVIII-11.

Midline parieto-occipital EEG and tremograms for a patient (E. C.) with a pronounced unilateral rhythmic tremor of the outstretched right upper extremity. Apparent tremor recorded from the accelerometer on the left side represents purely transmitted motion from the right side. For this recording, the upper frequency response of the system was limited to 15 cps (1/2 amplitude) in order to eliminate a prominent EMG content in the EEG leads.
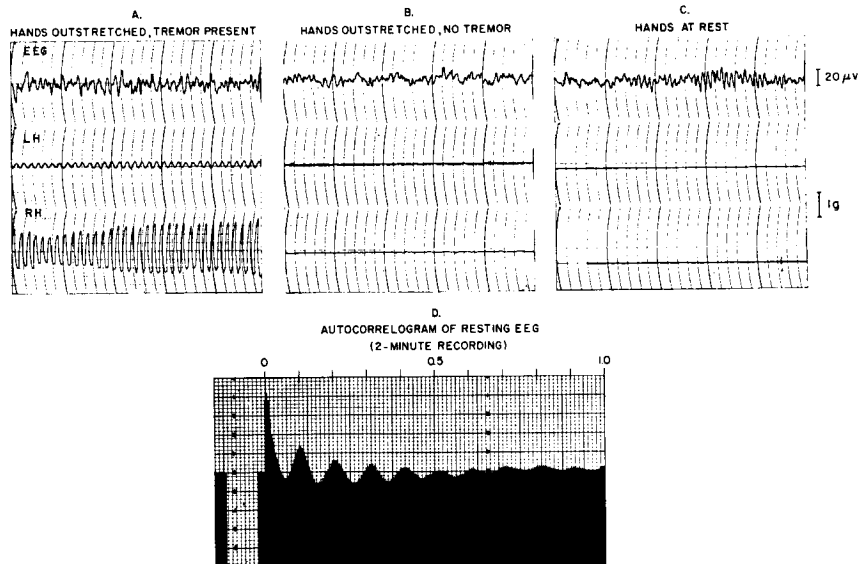
Fig. XVIII-12. Portions of EEG and tremogram for Patient E.C. for various conditions. (A). Both upper extremities in outstretched position. (B). Tracing obtained just after extremities are outstretched but before the rhythmical tremor had appeared. (C). Upper extremities at rest on the knees. (D). Autocorrelogram of resting EEG.

Recordings are shown in Fig. XVIII-11 from a patient with a pronounced rhythmical tremor that appeared upon sustained posturing of the right upper extremity in the out-stretched position. No tremor was present at rest on either side, and since the tremor appeared only the right side upon extending the limbs individually, the output of the accel-erometer on the left side in Fig. XVIII-11 must be considered to have resulted from mechanical crosscoupling, through the trunk of the body, from the tremor on the right side. The average tremor frequency is seen from the autocorrelogram to be 6/sec. A very prominent 6/sec component is also present in the autocorrelogram of the EEG, and a crosscorrelation coefficient of 0.35 is apparent from the crosscorrelogram for this frequency component in the EEG and in the right-sided tremor. Identical findings resulted from a second recording. The question of the possibility of an artefact of move-ment of the EEG electrodes or leads, as a source of the 6/sec component in the EEG trace must, of course, be immediately raised. A third accelerometer was not available to permit direct monitoring of motion of the head, but inspection of the ink traces shown in Fig. XVIII-12 suggests that the 6/sec component in the EEG ink trace did not derive from movement artefact. Thus, upon close inspection of the EEG recorded when the upper extremities are in the outstretched position (A, in Fig. XVIII-12) a 6/sec com-ponent is evident, but its amplitude does not appear to parallel that of the tremor, as would be expected if the former were purely a movement artefact. Moreover, a 6/sec component is evident in the EEG in the few seconds just after the upper limbs are out-stretched, but before the tremor has developed, as is apparent in B, in Fig. XVIII-12. The EEG in the latter instance contrasts with that recorded when the upper extremities are completely at rest (C, in Fig. XVIII-12), which is characterized by rather prominent alpha activity of an average frequency of 10/sec, as is evident from the autocorrelogram shown in D in Fig. XVIII-12. Hence it appears unlikely that movement artefact is a source of the 6/sec component in the EEG in Fig. XVIII-11, but this possibility cannot be excluded entirely.

Recordings from a patient with rather rhythmic tremors of the physiological type bilaterally are shown in Figs. XVIII-13 through XVIII-19. In Fig. XVIII-13, the tremo-grams and the parieto-occipital EEG tracings are very similar, but no such similarity is apparent between the tremograms and the central EEG leads. The central EEG leads were included in order to explore possible relationships between the tremor and the EEG overlying the pre-central motor area. The similarity of the autocorrelograms of the tremor recordings and those of the parieto-occipital EEG recordings is also evident from inspection of Fig. XVIII-14. The average frequency of the tremor on the left side is approximately 8.5/sec, on the right it is slightly lower, approximately 8/sec. The average frequency in the parieto-occipital EEG is approximately 9.5/sec; no prominent rhythmic activity is present from the central EEG leads, as evidenced by the auto-correlograms for the latter.
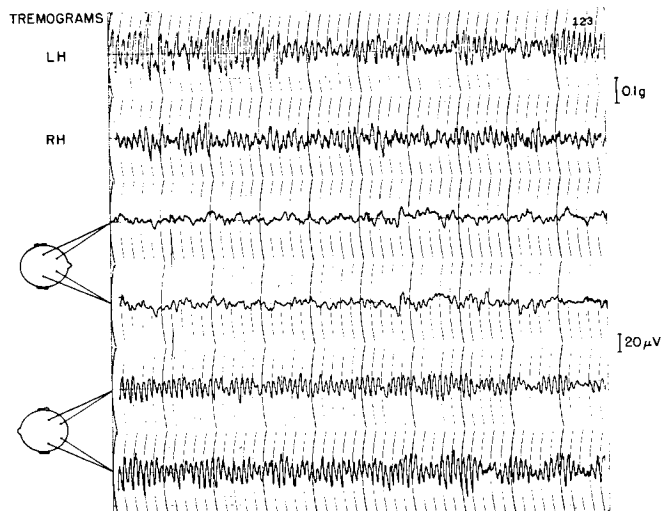
Fig. XVIII-13. Bilateral tremograms and EEG recordings from bilateral fronto-central and parieto-occipital electrodes for a patient (P. D.) with rather rhythmic physiological tremors bilaterally.
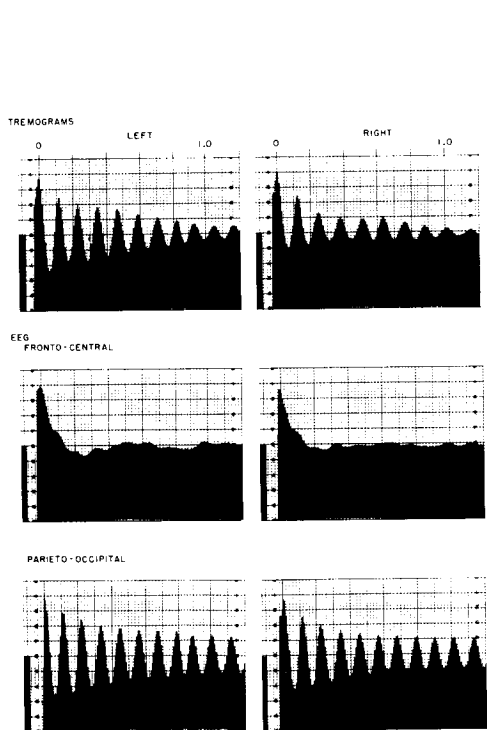
Fig. XVIII-14.

Autocorrelograms of tremograms and EEG for Patient P. D.
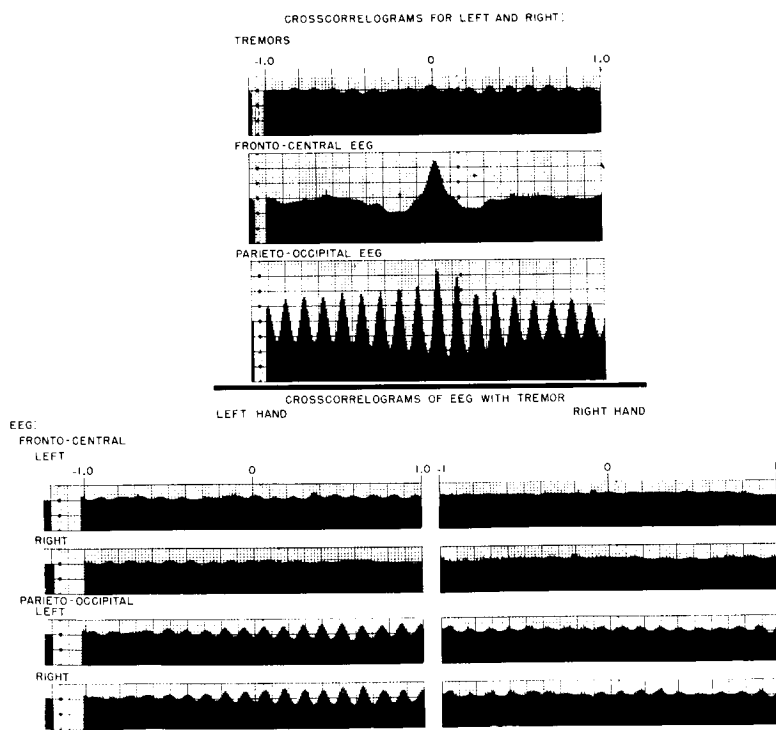
Fig. XVIII-15.

Crosscorrelograms of EEG with tremograms for Patient P. D.
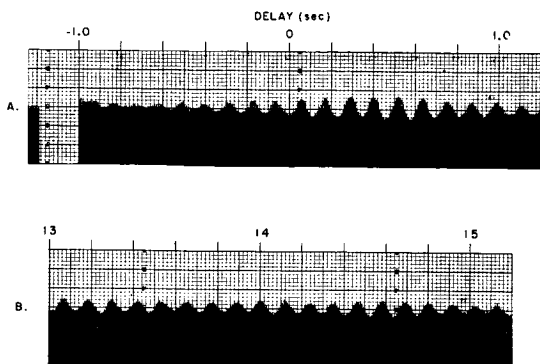
DELAY (sec)

Fig. XVIII-16.

Crosscorrelogram of left parieto-occipital EEG with tremogram of left upper extremity: (A) for delays in the range −1 - +1 sec; (B) for delays in the range +13 - +15 sec. Duration of recordings analyzed, 2 minutes. (Patient P. D.)

FLASH RATE (per sec)
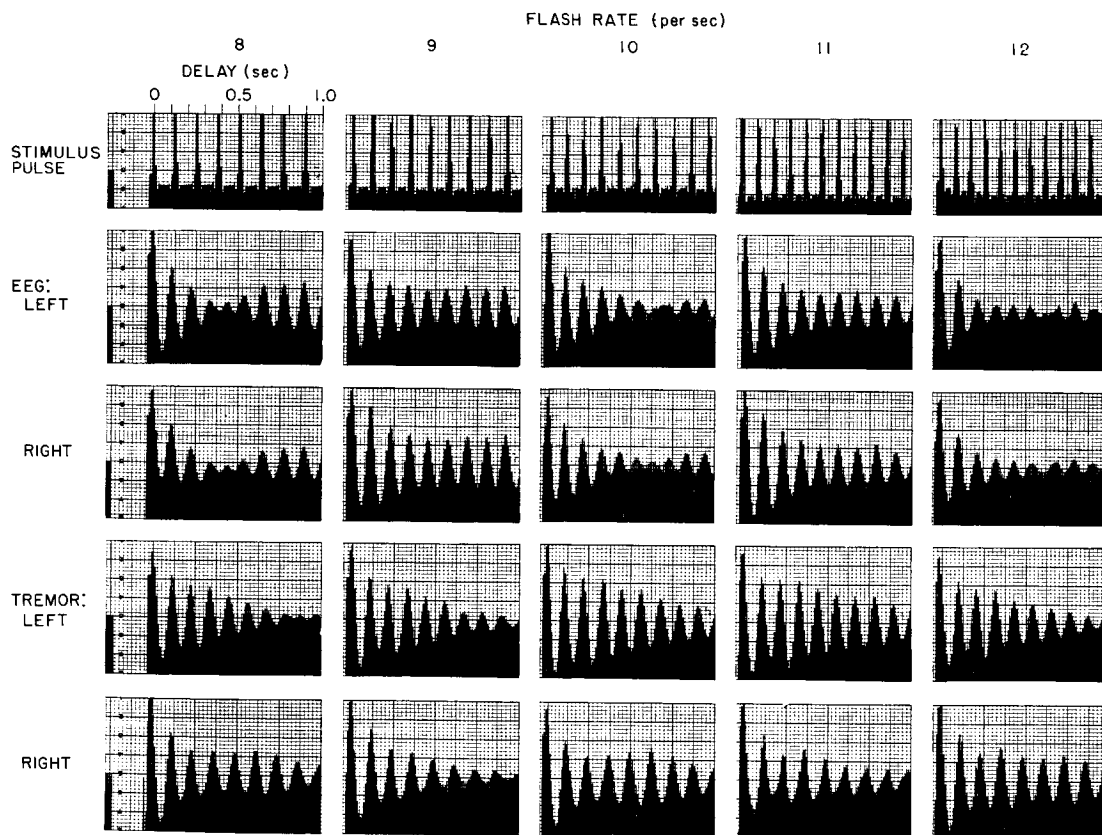
Fig. XVIII-17. Autocorrelograms of parieto-occipital EEG and tremograms for Patient P. D. for various flash rates of photic stimulation. Duration of recordings, 40 sec.

Crosscorrelograms of the tremor recordings on the two sides with the several EEG leads are shown in Fig. XVIII-15, and it is apparent that the amplitudes of none of the crosscorrelograms exceed the level of sampling artefact, with the possible exception of the correlograms for the tremor on the left with the parieto-occipital EEG, a finding that may be of interest in view of the fact that this patient was left-handed. Since the parieto-occipital EEG on the two sides are strongly correlated, as shown in the lower correlogram in Fig. XVIII-15, it is not surprising that the crosscorrelograms of the tremor on the left side with the parieto-occipital EEG from both sides are quite similar.

In view of the very similar autocorrelograms for the tremor on the left with the parieto-occipital EEG (Fig. XVIII-14), the question arose as to whether the apparent crosscorrelation of these two resulted merely as an artefact of sampling in this 2-minute recording, particularly in view of the fact that the peak in the crosscorrelograms is at a delay of ~0.5 sec, rather than nearer zero delay. To examine this possibility, the crosscorrelogram was recomputed for the full range of delays extending from -1 to +15 sec (the upper limit available with the correlator), under the assumption that any apparent correlation remaining after such large delays would be due solely to sampling artefact. The initial and final segments of this continuously computed curve are shown in Fig. XVIII-16. Following the peak in the crosscorrelogram at 0.62 sec at no intermediate delay did the amplitude exceed the value apparent in Fig. XVIII-16 at 14.6 sec delay, that is, for no delays following the former did the amplitude of the crosscorrelogram exceed more than 0.6 its value at 0.62 sec, although, at several intermediate delays, the amplitude was approximately the same as that at 14.6 sec. The last may then be taken as an indication of the peak level of spurious correlation caused by sampling artefact, and it is clearly exceeded for delays in the vicinity of 0.6 sec. (As a check on the performance of the over-all system for correlation at such large delays, a 10/sec sine wave was recorded on the magnetic tape and processed in exactly the same manner as for the crosscorrelogram in Fig. XVIII-16, the crosscorrelogram being computed out to a delay of 15 sec (150 wavelengths). The amplitude of the envelope of the correlogram at a maximum delay of 15 sec had decreased by only 14 per cent of amplitude at zero delay, the decrease being due to slight irregularities in the speed of the magnetic drum and of the magnetic tape recorders.) It is apparent, then, that sampling artefact does not appear to account for the crosscorrelation in the range of ~0.6 sec delay between the left-sided tremor and the parieto-occipital EEG. The possibility that movement artefact in the EEG could be the basis for the apparent correlation between the left-sided tremor and the EEG would appear to be excluded by the absence of correlation of the left-sided tremor with the EEG from the central leads, and of the right-sided tremor (whose amplitude was approximately the same as that on the left) with any of the EEG leads.

Correlograms for this patient from recordings with repetitive photic stimulation are

shown in Figs. XVIII-17 through XVIII-19. From the autocorrelograms of the tremor on the left and right sides, as shown in Fig. XVIII-17, it is apparent that the average frequencies of the tremor on the two sides remain virtually unchanged for different flash rates, the frequencies being only slightly different from those with no flash (Fig. XVIII-14). At the same time, from Fig. XVIII-17, it is evident that the dominant rhythmic activity in the parieto-occipital EEG is little altered by the flashing. As was the case with no photic stimulation (Fig. XVIII-14) the average frequency of the tremor on the left side is very similar to that of the parieto-occipital EEG activity, of 9/sec. The tremor rate on the right side remained at 8/sec.

For a closer examination of the question of relationships between tremor and EEG, the crosscorrelograms shown in Figs. XVIII-18 and XVIII-19 were computed. Upon first sight, the crosscorrelograms of the tremor on the left side with the parieto-occipital EEG activity (Fig. XVIII-18) for these very brief (40-sec) recordings appear to suggest that these two phenomena are in fact correlated. Upon closer inspection of the cross-correlograms, however, it becomes apparent that the location of the peaks nearest zero delay are not consistently located on the delay scale; they appear in fact to be randomly distributed about zero delay. Hence we must conclude that the apparent crosscorrelation of tremor with EEG activity represents entirely an artefact of sampling in these short recordings. The same is true for the crosscorrelograms of tremor on the right side with EEG activity, which are shown in Fig. XVIII-19, the amplitudes of the crosscorrelograms being generally smaller than those in Fig. XVIII-18. From these results, it is clear that the tremor in this patient was not entrained by the photic stimulation, nor was the tremor synchronized with the parieto-occipital EEG during the stroboscopic stimulation.

To summarize, correlation analysis of recordings from two normal controls has not indicated that the accelerometrically monitored physiological tremors from the outstretched upper extremities are correlated with the parieto-occipital EEG. There is evidence that such a relationship was present in recordings from a patient with a pronounced unilateral rhythmic tremor, but the possibility that this apparent correlation may have arisen from a component of movement artefact in the EEG has not been completely excluded. Recordings from a left-handed patient with rather rhythmic tremors of the physiological type bilaterally were suggestive of a crosscorrelation between the tremor on the left side with the parieto-occipital EEG. Such a correlation was not distinguishable from that caused by sampling artefact in shorter recordings (40 sec instead of 120 sec) with photic stimulation at flash rate in the alpha-frequency range. The latter stimulation at several flash rates was without effect upon the frequency of the tremor. Similarly, no effect of intermittent photic stimulation upon tremor was found for one of the normal control subjects.

The fact that the frequency of physiological tremors, the significance of which is
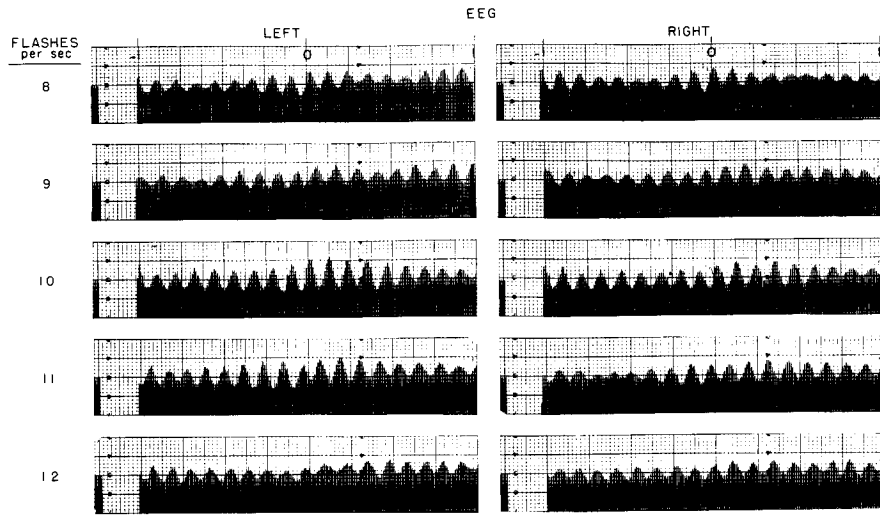
Fig. XVIII-18. Crosscorrelograms of parieto-occipital EEG with tremograms of left side for various flash rates for Patient P. D. Duration of recordings, 40 sec.
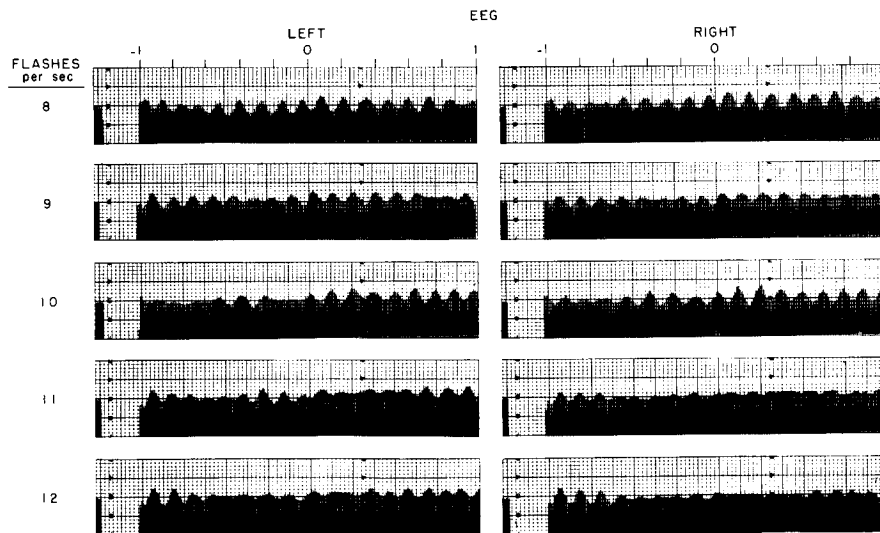
Fig. XVIII-19. Crosscorrelograms of parieto-occipital EEG with tremograms of right side for various flash rates for Patient P. D. Duration of recordings, 40 sec.

still under discussion,[2,3] may lie within the alpha-frequency range of the EEG has long been of interest.[4-13] (Despite the fact that in some instances the characteristics of the monitoring system for the tremors may influence the frequency of the recorded tremor,[14,15] the fact that tremors have been recorded with several different techniques, some of which (for example, photoelectric) cannot alter the frequency characteristics of the tremor itself, indicates that the basic phenomenon cannot be an artefact of the recording system.) There has been no agreement, however, concerning the question of whether physiological tremor activity is consistently correlated with EEG activity. Thus, in some instances, short-time wave-to-wave correspondence,[4-7] as well as statistical agreement between the mean frequencies of the two phenomena[9] (as determined by visual analysis), have been reported, but in other series[13] no such agreement has been found.

It seems clear that physiological tremor is not solely dependent upon rhythmic activity of the cerebral cortex, for the former has been observed when the latter is not present, or when the connections between brain and limb, at the level of the spinal cord, are interrupted.[7,8] The possibility that components in both, under particular circumstances, could be entrained by a common pacemaker at such a subtle level that no wave-to-wave correspondence would be evident upon inspection of the simultaneously recorded inked traces, might be considered. Such records might or might not exhibit the same average frequency, as determined by visual or automatic frequency analysis. By cross-correlation analysis, however, the presence of such a common (entrained) component could be detected, even if it were not at all evident in the original traces.[19]

Alternatively, for two quasi-rhythmic systems possessing similar or even identical frequency characteristics which in reality are quite independent of one another, a wave-to-wave correspondence, of an entirely fortuituous nature, could appear for brief intervals of time. Such true independence could be demonstrated with the aid of crosscorrelation analysis, appropriate attention being paid to the possibility of correlation errors arising from sampling artefact.[20]

The present study, which demonstrates the particular suitability of correlation analysis for study of tremor-EEG relationships, has not unequivocally indicated the existence of a consistent relationship between physiological tremor and the EEG, within the limitations of the present experimental conditions, for the subjects who were examined. Further studies of the problem, including simulation ones, are planned.

J. S. Barlow

References

1. J. S. Barlow, Crosscorrelation of accelerometric recordings of movement disorders in man, Quarterly Progress Report 76, Research Laboratory of Electronics, M. I. T., January 15, 1965, pp. 293-311.

2. M. Haider and D. B. Lindsley, Microvibrations in man and dolphin, Science 146, 1181-1183 (1964).

3. H. Rohracher, Microvibration, permanent muscle-activity, and constancy of body-temperature, Perceptual and Motor Skills 19, 198 (1964).

4. H. H. Jasper and H. L. Andrews, Brain potentials and voluntary muscle activity in man, J. Neurophysiol. 1, 87-100 (1938).

5. H. H. Jasper, Cortical excitatory states and synchronism in the control of bioelectric autonomous rhythms, Cold Spring Harbor Symposia on Quantitative Biology 4, 320-338 (1936).

6. R. W. Lansing, Relation of brain and tremor rhythms to visual reaction time, EEG Clin. Neurophysiol. 9, 497-504 (1957).

7. T. Lindqvist, Finger tremor and the α-waves of the electroencephalogram, Acta Med. Scand. 108, 580-585 (1941).

8. J. Marshall and E. G. Walsh, Physiological tremor, J. Neurol. Neurosurg. Psychiat. 19, 260-267 (1956).

9. O. Nirkko, On the individual correspondence between the EEG alpha rhythm and muscle vibration, Report No. 2, The Psychological Institute, University of Helsinki, 1961.

10. H. Rohracher, Alpha-Wellen und Mikro-Schwingungen, Acta Neurochirurgica 3 (Suppl.), 369-370 (1954).

11. H. Rohracher, Mechanische Mikroschwingungen des menschlichen Körpers (Urban U. Schwarzenberg, Vienna, 1949), pp. 17-18.

12. H. Sugano and K. Inanaga, Studies on minor tremor, Jap. J. Physiol 10, 246-257 (1960).

13. L. T. Travis and C. N. Cofer, The temporal relationship between brain potentials and certain neuromuscular rhythms, J. Exper. Psychol. 21, 565-569 (1937).

14. A. M. Hamoen, On the physiology of tremor, EEG Clin. Neurophysiol. 10, 752 (1958).

15. J. G. L. Williams, A resonance theory of "microvibrations," Psychol. Rev. 70, 547-558 (1963).

16. J. S. Barlow, Autocorrelation and crosscorrelation analysis in electroencephalography, IRE Trans. Vol. ME-6, pp. 179-183, 1959.

17. W. A. Rosenblith (ed.), Processing Neuroelectric Data, The M. I. T. Press, Cambridge, Mass., 2d printing, 1962, 127 pp.

18. W. B. Davenport, Correlator errors due to finite observation intervals, Technical Report 191, Research Laboratory of Electronics, M. I. T., March 8, 1951.

19. Compare Fig. 4 in J. S. Barlow[16] and Fig. 3. 17 of W. A. Rosenblith (ed.)[17], p. 54.

20. Compare W. B. Davenport,[18] and W. A. Rosenblith (ed.)[17], pp. 56-60.

E. ACTIVE AND COMPENSATORY VISUAL-OCULOMOTOR TRACKING IN RELATION TO THE VESTIBULAR SYSTEM

In earlier experiments[1-3] designed to study quantitatively the nature of the disorder of stabilization of the visual image which follows complete loss of vestibular function in man, it was found, for one such patient, that the compensatory oculomotor tracking (fixation point stationary, head rotated passively while the subject was seated in the motionless chair) was indistinguishable from compensatory oculomotor tracking by a normal subject. The possibility that neck proprioceptive mechanisms might provide the sensory basis for the remarkably accurate passive oculomotor tracking by the patient was considered, but this possibility could not be directly tested experimentally in those experiments.

To further explore the nature of visual-oculomotor tracking in relation to vestibular and neck proprioceptive mechanisms, the present series of experiments, employing a Bárány chair, were carried out. For this purpose, a Bárány chair was modified so that its angular position in the horizontal plane could be monitored electrically by means of a potentiometer whose wiper is driven by the pivot of the chair. The electro-oculographic (EOG) recording technique and the use of correlation techniques for analyses of the data have been described previously.[1] For the present recordings, which were limited to study of eye movements in the horizontal, one electrode was placed adjacent to the outer canthus of each eye.

Experiments were carried out with two patients who had suffered loss of vestibular function following streptomycin toxicity to the labyrinthine apparatus, and with one neurological patient with normal vestibular function and two normal adults who served as controls. The recordings from one of the patients with vestibular loss were technically unsatisfactory for analysis (because of poor cooperation by the patient). Results from the other patient with this disorder and from the control subjects are described below.

Recordings were carried out as follows:

1. Compensatory visual oculomotor tracking, eyes open.

2. Compensatory oculomotor tracking, eyes open, with proprioceptive input: same as (1), except that the subject maintains fixation on his thumb, which overlies the spot on the oscilloscope, the hand grasping the front of the oscilloscope.

3. Compensatory tracking, eyes closed: the subject is instructed to attempt to visualize and track the stationary object of (1) above, as he is moved passively in the Bárány chair.

4. Compensatory tracking, eyes closed, with proprioceptive input: same as (3), but with proprioceptive input as in (2).

5. Active tracking (eyes open): the subject attempts to track the spot of light on

the oscilloscope as it moves randomly along the x-axis of the oscilloscope screen. (Input signal to the x-amplifier of the oscilloscope: filtered noise in the band with DC to 0.8 cps with an attenuation of 36 db per octave for higher frequencies.)

Results for the two young normal controls are shown in Figs. XVIII-20 through XVIII-28. In Fig. XVIII-20, ink tracings and correlograms for compensatory visual tracking are shown for subject T. D. The electrical positions of eye and chair are remarkably similar, as are the two autocorrelograms. In the crosscorrelogram, the appearance of the peak at zero delay indicates that there is no time lag between eye position and chair position, the former compensating exactly for the latter, as the gaze is maintained at the stationary object. No change is apparent upon the addition of proprioceptive input which is afforded as the subject fixates on his stationary thumb, as is apparent from comparison of Fig. XVIII-21 with Fig. XVIII-20.

In Fig. XVIII-22 results are shown for the recording in which tracking of the stationary object was attempted (by imagining the fixation point) while the eyes were closed. Although it is evident from inspection of the ink tracings that there is some tendency for eye movements to parallel those of the chair, the EOG is much more irregular in comparison with the tracing of the chair position than was the case in Fig. XVIII-20 with the eyes open. Moreover, it is apparent from the crosscorrelogram that eye position lagged the chair position by approximately 50 msec. The addition of proprioceptive information while the eyes were closed (Fig. XVIII-23) was ineffective in improving the quality of the attempted tracking. The time lag in the latter instance of eye position with respect to chair position was approximately 60 msec.

For comparison with the results of compensatory oculomotor tracking shown in Fig. XVIII-20, the findings for active tracking of the horizontally moving object (the spot of light on the oscilloscope screen, the subject remaining stationary in the Bárány chair) are shown in Fig. XVIII-24. The irregular EOG tracing which results from successive saccadic jumps is in striking contrast to the completely smooth EOG for compensatory tracking shown in Fig. XVIII-20. The much greater irregularity of the eye movements in Fig. XVIII-24 cannot be accounted for solely by the fact that the average frequency of the tracking signal in the latter instance was somewhat higher than that used in Fig. XVIII-20, inasmuch as saccadic movements are apparent even for the lowest frequency components in Fig. XVIII-24. Despite this irregularity, however, the average lag of eye movement with respect to spot movement is only 8 msec, as is indicated by the location of the peak in the crosscorrelogram in Fig. XVIII-24.

Results of similar recordings from another normal subject are shown in Figs. XVIII-25 through XVIII-28. In Fig. XVIII-25 are shown the findings for compensatory tracking with eyes open, and it is evident that the EOG tracing is an almost exact replica of the electrical recording of the chair position. The addition of proprioceptive input resulted in no change in the precision of oculomotor tracking.
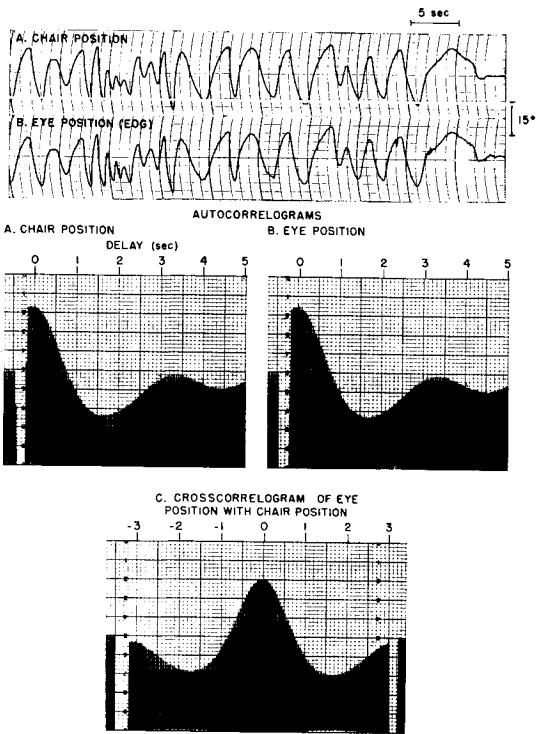
Fig. XVIII-20.

Compensatory visual-oculomotor tracking of a stationary object (spot of light on an oscilloscope screen). Subject seated in Bárány chair whose angular position is monitored electrically. Eye position monitored electrooculographically by means of electrodes at the outer canthus of each eye. The angular calibration is the same for Figs. XVIII-20 through XVIII-33. (Note that a different angular calibration is employed for Figs. XVIII-34 through XVIII-38.) Duration of recording analyzed, 4 minutes. Delay scale in seconds. (Subject T. D.)
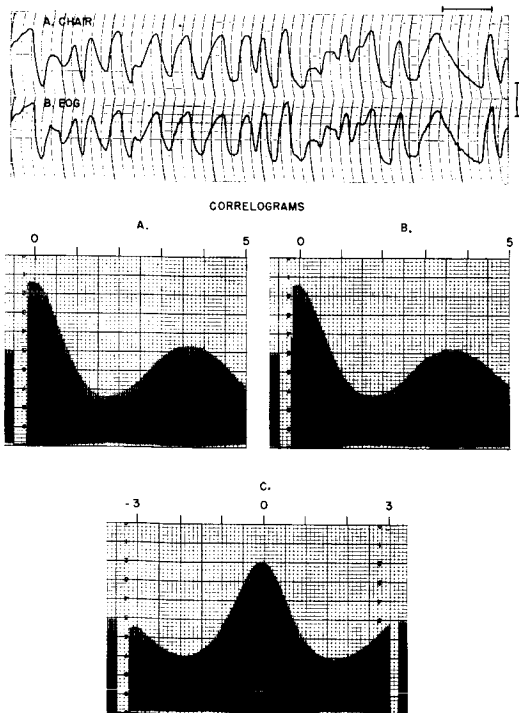
Fig. XVIII-21. Compensatory tracking, eyes open, with proprioceptive input. (Subject T. D.)
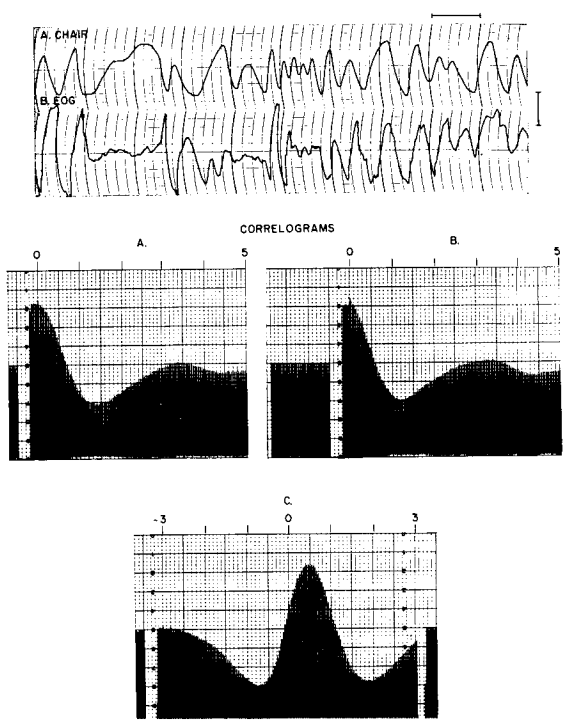
Fig. XVIII-22.

Attempted compensatory tracking with eyes closed. (Subject T.D.)
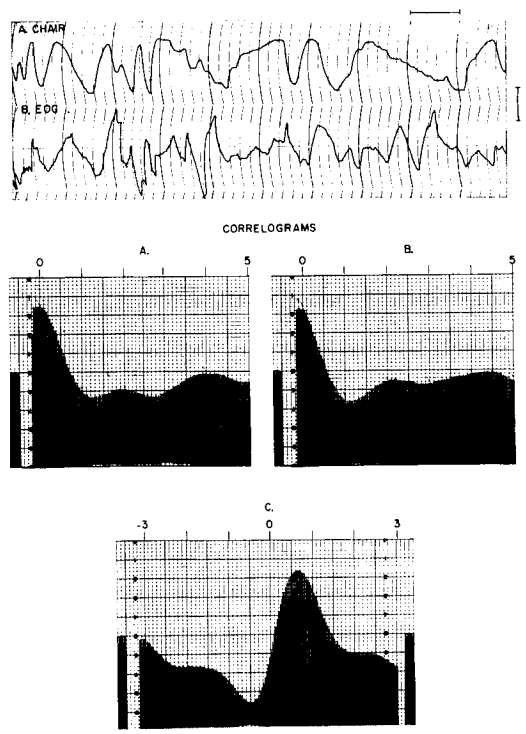
Fig. XVIII-23.

Attempted compensatory tracking, eyes closed, with proprioceptive input. (Subject T.D.)
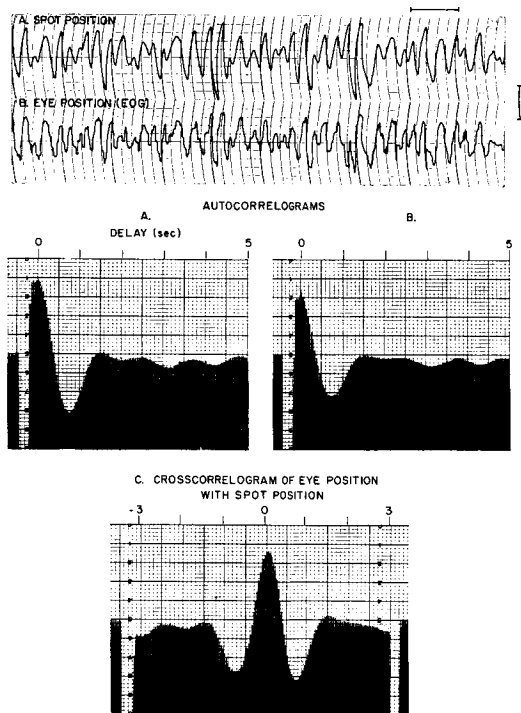
Fig. XVIII-24. Active tracking of a horizontally moving object (spot of light on an oscilloscope screen, X-input to oscilloscope, filtered noise, DC-0.8 cps). Subject seated motionless in Bárány chair. (Subject T.D.)
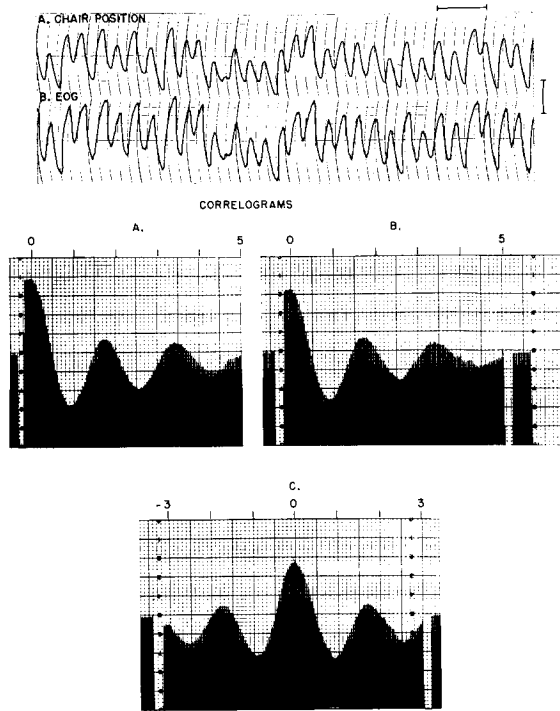
Fig. XVIII-25. Compensatory tracking, eyes open. (Subject C. M.)
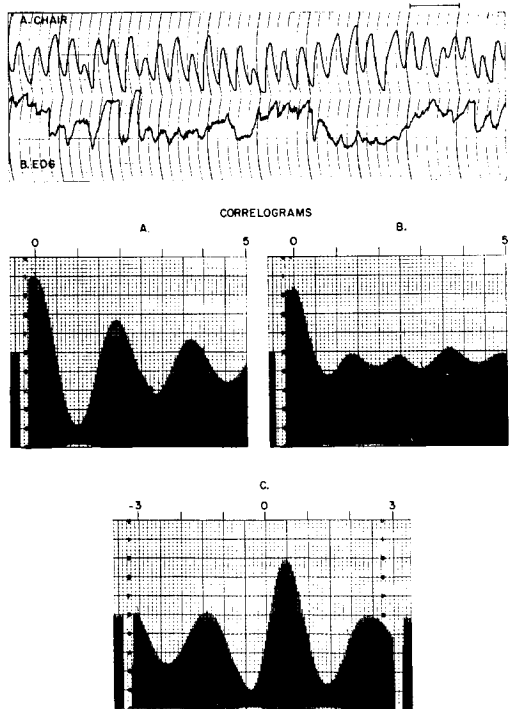
Fig. XVIII-26. Attempted compensatory tracking with
eyes closed. (Subject C. M.)
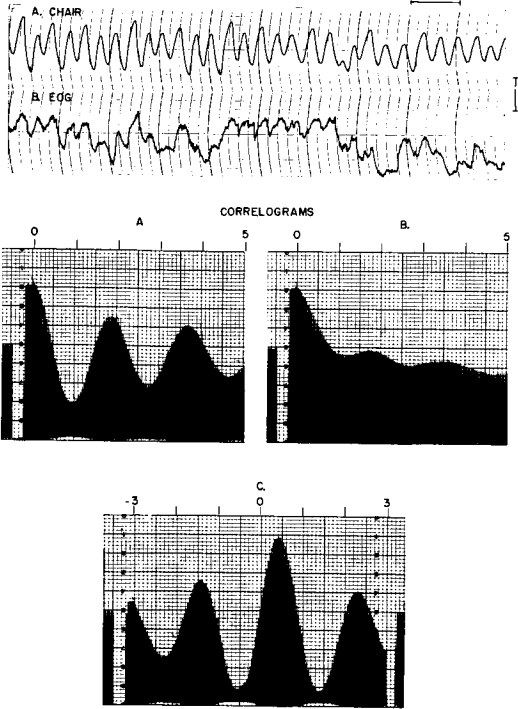
Fig. XVIII-27.

Attempted compensatory tracking, eyes closed, with proprioceptive input. (Subject C. M.)
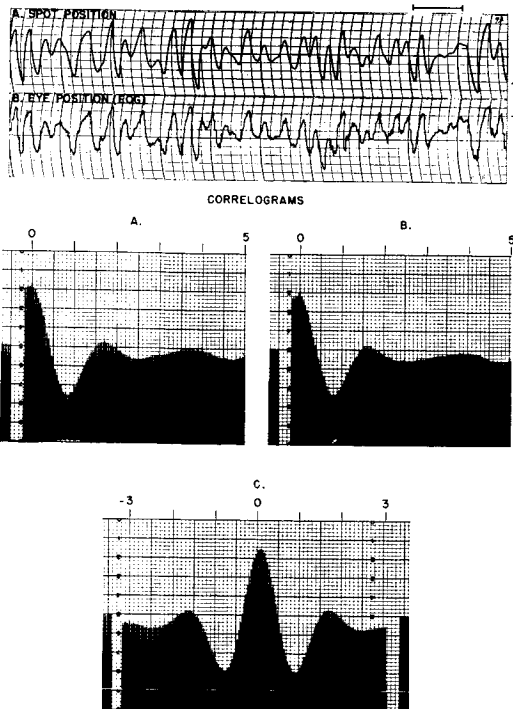
Fig. XVIII-28.

Active tracking of horizontally moving target, eyes open, subject stationary. (See Fig. XVIII-24.) (Subject C. M.)

Attempted tracking of the imagined fixation point, while the eyes were closed, results in an EOG tracing that bears little resemblance to the tracing of chair position (Fig. XVIII-26). There does, however, remain some correlation between the two, as is indicated by the crosscorrelogram, which shows that for the component common to the two, the eye position lagged chair position by ~ 500 msec on the average. As is evident from Fig. XVIII-27, the addition of proprioceptive input results in little change.

Findings for active tracking of the moving spot by the stationary subject (Fig. XVIII-28) indicate the presence of prominent saccadic components in the EOG, and the crosscorrelogram indicates an average lag time of approximately 100 msec.

Results of recordings from a patient with motor neuropathy, but with a normally functioning vestibular system, are shown in Figs. XVIII-29 through XVIII-33. Compensatory tracking with eyes open by this patient (Fig. XVIII-29) is much the same as that for the normal subjects shown in Figs. XVIII-20 and XVIII-25, and no change resulted when proprioceptive input was added (Fig. XVIII-30). Upon closing of the eyes, however, there was almost complete "decoupling" between the EOG and the electrical signal of chair position, as is evident from Fig. XVIII-31, a finding that remained the same when proprioceptive input was added while the eyes were closed (Fig. XVIII-32). It is of interest that this patient was unable to carry out active tracking of the tracking signal that was employed in these recordings (Fig. XVIII-33), a finding that contrasts strikingly with the excellent compensatory tracking (Fig. XVIII-29).

Results of recordings from a patient who had lost vestibular function from streptomycin toxicity to the vestibular apparatus approximately one year previously are shown in Figs. XVIII-34 through XVIII-38. The absence of vestibular function for this patient was evident from the fact that there was no response with electro-oculographic monitoring to air calorics, iced alcohol calorics, nor any nystagmus on positional testing. The extent to which the EOG tracing for this patient resembles that of the electrical indication of chair position, in passive oculomotor tracking, as shown in Fig. XVIII-34, is truly remarkable. Provision of proprioceptive input results in no change (Fig. XVIII-35); the time lag between the EOG and the chair-position signal in both instances is zero, as is indicated by the location of the peak in the crosscorrelograms in Figs. XVIII-34 and XVIII-35. Compensatory tracking attempted when the eyes are closed (Fig. XVIII-36) indicates that the degree of "coupling" between the eye position and chair position is considerably less than when the eyes are open, although it is not completely absent in the former instance. It is of interest that the crosscorrelogram of Fig. XVIII-36 indicates that the component of the EOG which was correlated with the chair position leads the latter by approximately 120 msec. Provision of proprioceptive input with eyes closed resulted in no essential change (Fig. XVIII-37).

Active visual oculomotor tracking by this patient (Fig. XVIII-38) is remarkably smooth; in fact, it is better than that for any of the control subjects (Figs. XVIII-24,

A. CHAIR POSITION

B. EOG

AUTOCORRELOGRAMS

A.

B.

C.
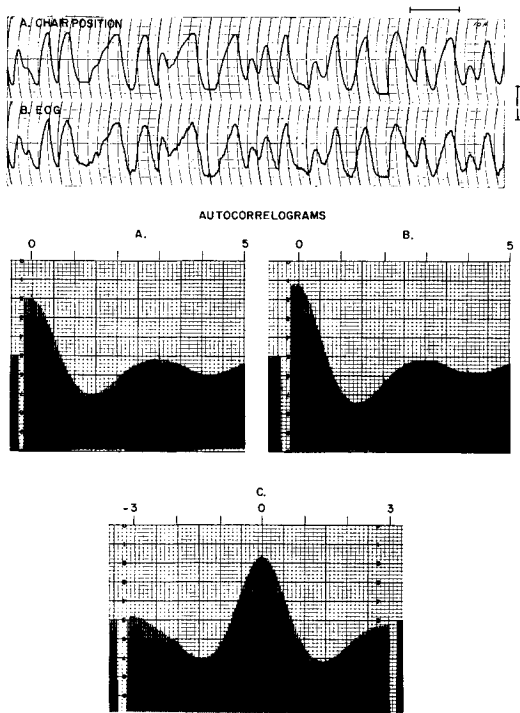
Fig. XVIII-29.   Compensatory tracking, eyes open.   (Patient C. E.)
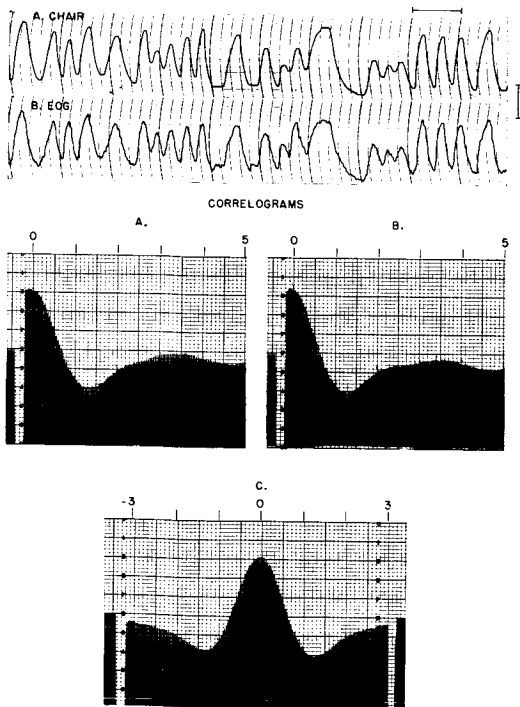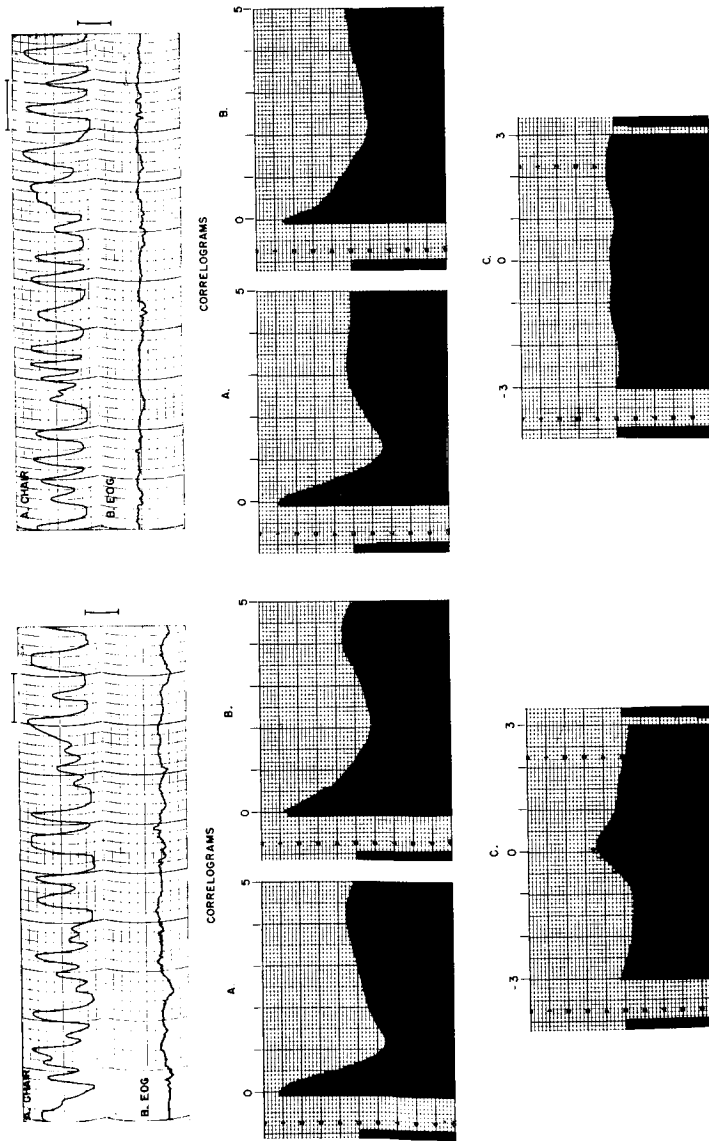
A. CHAIR

B. EOG

CORRELOGRAMS

A.

B.

C.

Fig. XVIII-30.

Compensatory tracking, eyes open, with proprioceptive input.   (Patient C. E.)

Fig. XVIII-33.

Attempted active tracking, eyes open, of a horizontally moving target by Patient C. E. The patient was unable to execute this maneuver; hence, no correlograms were computed.
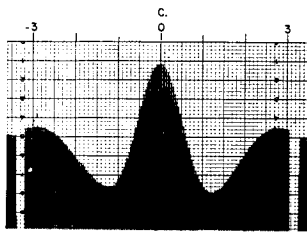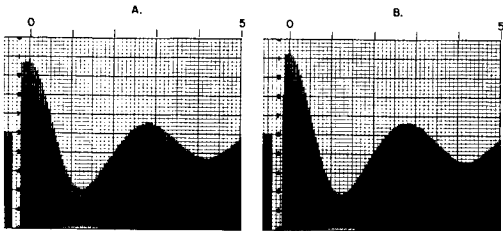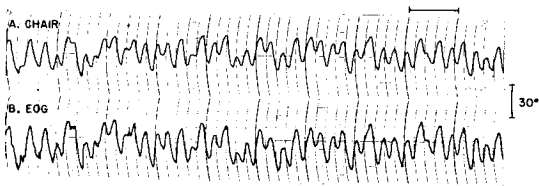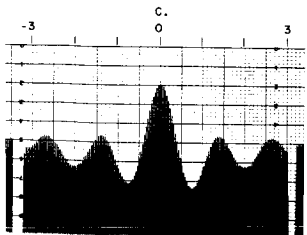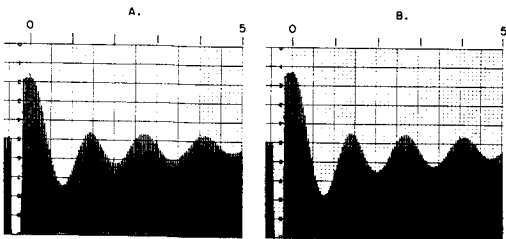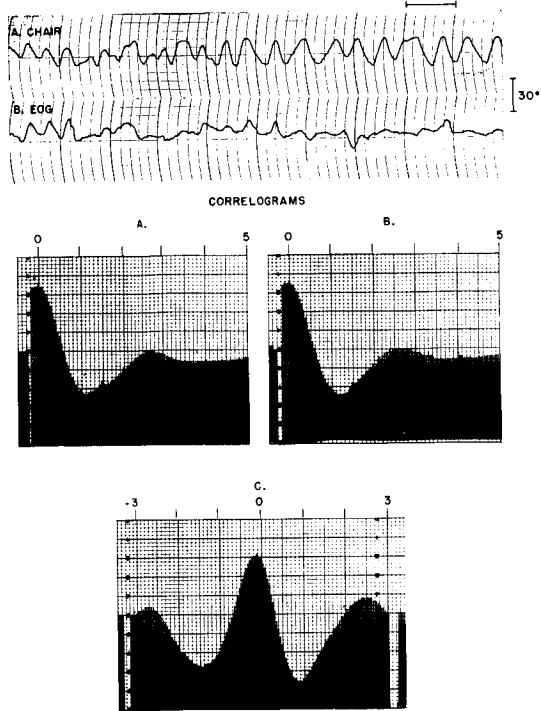

Fig. XVIII-32.

Attempted compensatory tracking, eyes closed, with proprioceptive input. (Patient C. E.)

Fig. XVIII-31.

Attempted compensatory tracking, eyes closed. (Patient C. E.)
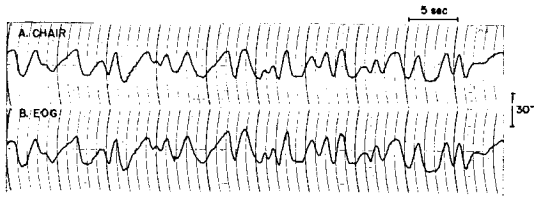
Fig. XVIII-34. Compensatory tracking, eyes open. (Patient K. H.)

Fig. XVIII-35. Compensatory tracking, with proprioceptive input, eyes open. (Patient K. H.)

Fig. XVIII-36. Attempted compensatory tracking, eyes closed. (Patient K. H.)
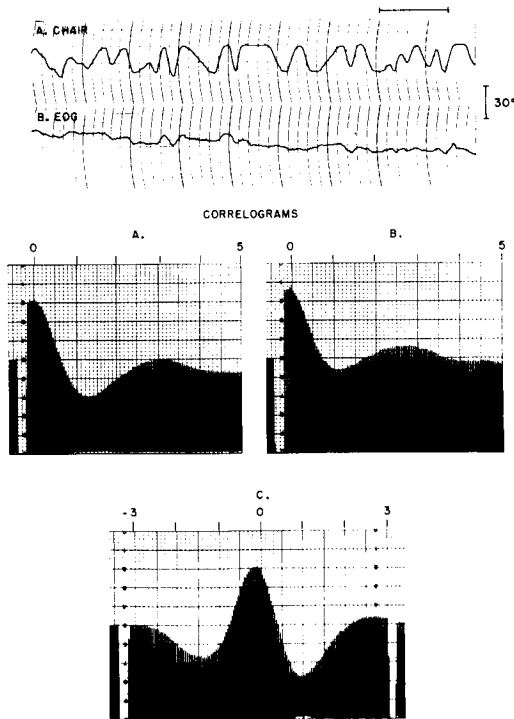
Fig. XVIII-37. Attempted compensatory tracking, eyes closed, with proprioceptive input. (Patient K. H.)

XVIII-28 and XVIII-33). Moreover, there is no time lag, on the average, between EOG and the position of the spot (the tracking signal) as evidenced by the location of the peak in the crosscorrelogram in Fig. XVIII-38 at zero delay.

The findings from the present series of experiments, together with those previously reported,[1] provide further quantitative information concerning the nature of active and compensatory visual-oculomotor tracking, and suggest the following points in relation to individuals with a normally functioning vestibular system:

1. Compensatory oculomotor tracking (stationary fixation point, passively moving subject) is executed quite as well if the whole body is moved as a unit in the Bárány chair as it is when only the head is moved and the rest of the body remains stationary. It is therefore evident that neck proprioceptive mechanisms have no essential role in compensatory oculomotor tracking of this type.[4]

2. The "decoupling" of ocular movements which appears when the eyes are closed is quite pronounced both in movements of the head with respect to the neck,[5] and in movements of the entire body as a unit, while subjects are seated in a Bárány chair (Figs. XVIII-22, XVIII-26, and XVIII-31). Therefore, neck proprioceptive mechanisms appear to have no essential role in eye movements when the eyes are closed.

3. Compensatory oculomotor tracking in response to passive motion of the subject in a Bárány chair when the eyes are open, is equally good with or without the proprioceptive input that is provided by the stationary thumb as an object for fixation of gaze. (Compare Figs. XVIII-20 and XVIII-29 with Figs. XVIII-21 and XVIII-30.)

4. The extent of "decoupling" of ocular movements when the eyes are closed is quite pronounced with or without the above-mentioned type of proprioceptive input. (Compare Figs. XVIII-22, XVIII-26, and XVIII-31 with Figs. XVIII-23, XVIII-27, and XVIII-32.) It is therefore apparent that proprioceptive information from the upper limbs is not effective for the control of eye position when the eyes are closed.

5. Compensatory oculomotor tracking, carried out either during passive movement of the head alone, or during passive movement of the body as a unit, is much more smoothly accomplished than is active oculomotor tracking, the former being completely free of the saccadic jumps that characterize the latter. It is therefore evident that the fundamental neurophysiological bases of these two different kinds of tracking are different.

The following additional points are apparent from the results from the present and the previously reported findings in patients with vestibular loss:

6. Compensatory tracking either with motion of the head alone, or during motion of the body as a unit, is carried out quite normally in the absence of vestibular function.[6] Vestibular mechanisms, therefore, appear to have no essential role in the compensatory tracking observed in these experiments.

7. Active tracking was accomplished more smoothly by the present patient with
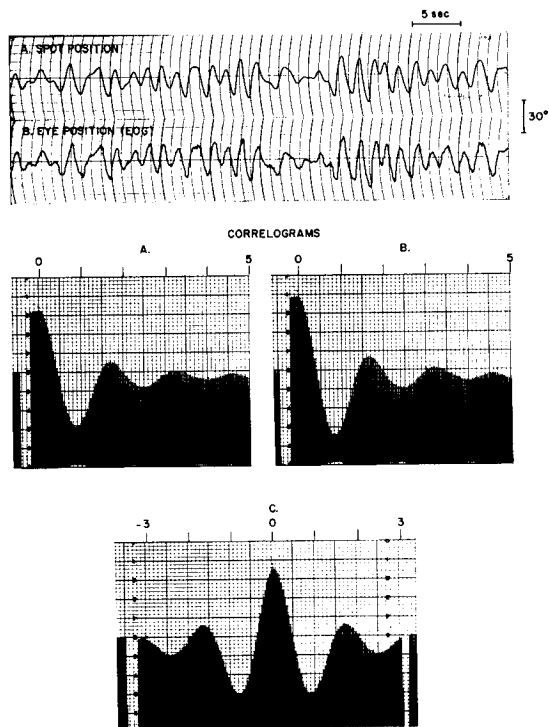
Fig. XVIII-38. Active tracking, eyes open, of a hori-
zontally moving target.   (Patient K. H.)

vestibular loss than by the controls, a finding that suggests a tendency to a compensatory effect for loss of vestibular function. The EOG tracings for active tracking by this patient (Fig. XVIII-38) are, however, still much less smooth than those for compensatory tracking (Fig. XVIII-34). It is apparent, then, that the physiological mechanisms for active tracking do not provide a basis for the smooth compensatory tracking that remains after vestibular loss in this patient.

These findings thus appear to indicate that the smooth compensatory visual-oculomotor tracking observed in these experiments is not based upon proprioceptive, nor vestibular mechanisms, nor upon the physiological mechanisms that are responsible for active or pursuit tracking. Accordingly, the physiological basis for compensatory tracking remains unclear. In view of the fact that the present patient had lost vestibular function approximately a year before these recordings, the possibility may exist that, in the intervening time, some as yet unelucidated mechanism may have compensated for the loss of vestibular function. It would therefore be of interest to make serial recordings from patients beginning soon after vestibular loss, in order to explore possible changes with time in the nature of compensatory tracking. Furthermore, in view of the fact that active tracking was carried out by this patient more smoothly than by the control subjects, it would be of interest to determine, by crosscorrelation analysis, oculomotor reaction times to step-function tracking signals.[7]

In future experiments, we plan to explore further the nature of the "decoupling" of ocular movements which occurs when the eyes are closed. As one step in the investigation of this phenomenon, recordings of compensatory tracking while the subject is in the dark and attempting to maintain gaze directed at an imagined stationary fixation point are planned, for comparison with recordings of compensatory tracking of a visible fixation point while the eyes are open, and for comparison with recordings of attempted compensatory tracking with the eyes closed.

J. S. Barlow, A. D. Weiss

(Dr. A. D. Weiss is Director of the Otoneurology Service, Massachusetts Eye and Ear Infirmary and Massachusetts General Hospital.)

References

1. J. S. Barlow, Control of eye movements in relation to the vestibular system and neck proprioceptive mechanisms in man, Quarterly Progress Report No. 65, Research Laboratory of Electronics, M.I.T., April 15, 1962, pp. 205-221.

2. J. S. Barlow, Study of control of eye movements in man by means of autocorrelation and crosscorrelation, Biometrics 18. 421, (1962).

3. J. S. Barlow, Analysis of Control of Eye Movements in Man by Means of Autocorrelation and Crosscorrelation, DIGEST of the 1962 15th Annual Conference of Engineering in Medicine and Biology, p. 4.

4. Compare Figs. XVIII-20, XVIII-25, and XVIII-29 with Fig. XVIII-8 of J. S. Barlow, Quarterly Progress Report No. 65, op. cit.

5. J. S. Barlow, Quarterly Progress Report No. 65, op. cit.; see Fig. XVIII-11.

6. Compare Fig. XVIII-34 with Fig. XVIII-15, J. S. Barlow, Quarterly Progress Report No. 65, op. cit.

7. J. S. Barlow, Quarterly Progress Report No. 65, op. cit., see Figs. XVIII-6 and XVIII-12.

XIX. NEUROPHYSIOLOGY[*]

W. S. McCulloch	B. Howland	Barbara C. Pickard
M. Blum	Susan Hurley	W. F. Pickard
J. E. Brown	K. Kornacker	W. H. Pitts
E. E. Fetz	J. Y. Lettvin	T. G. Smith
S. Frenk	Diane Major	A. Taub
R. C. Gesteland	T. McLardy	P. D. Wall
M. C. Goodall	L. M. Mendell	Barbara G. Wickelgren
	S. A. Papert	

A. NEW METHODS FOR TESTING CAMERA LENSES (Part II)

In Quarterly Progress Report No. 73 (pages 209-216) we described a new method for testing camera lenses and cameras. This method, which makes use of an especially designed analyzer lens placed in front of the camera lens, together with a polar coordinate test chart, permits simple diagnosis of the focal properties of the lens-camera combination. In particular, our method was shown to be sensitive to range-finder error, misalignment of lens-to-film plane, and to astigmatism and curvature of field.

Further tests with this method have indicated the need for the improved construction of the analyzer lens element which is reported here. Also, we evidently did not exhaust the potentialities of the method for the diagnosis of lens aberrations, since, with slight modification, our earlier work provides a sensitive test for spherical aberration and for coma. An additional, somewhat similar, test is also given which permits the determination of lateral chromatic aberration. Using both of these methods, we can determine five of the seven primary lens aberrations; the exceptions are distortion and axial chromatic aberration. Since these tests do not require the use of a precision optical bench, they should be particularly useful for amateur and professional photographers.

1. Improved Construction of the Analyzer Lens

In Fig. XIX-1 we show the basic method of lens testing. The analyzer lens CC is a crossed cylinder, i.e., a spherocylinder having equal positive and negative powers about two orthogonal axes, Y and Z. A narrow horizontal slit S is oriented at 45° to the axes of the crossed cylinder. For our previous tests we used ±1/8, ±1/4, and ±1/2 diopter ophthalmic cylinder lenses. Recent tests, however, indicate that the best available quality of such lenses of powers ±1/8 diopter or smaller give uncertain results in this application.
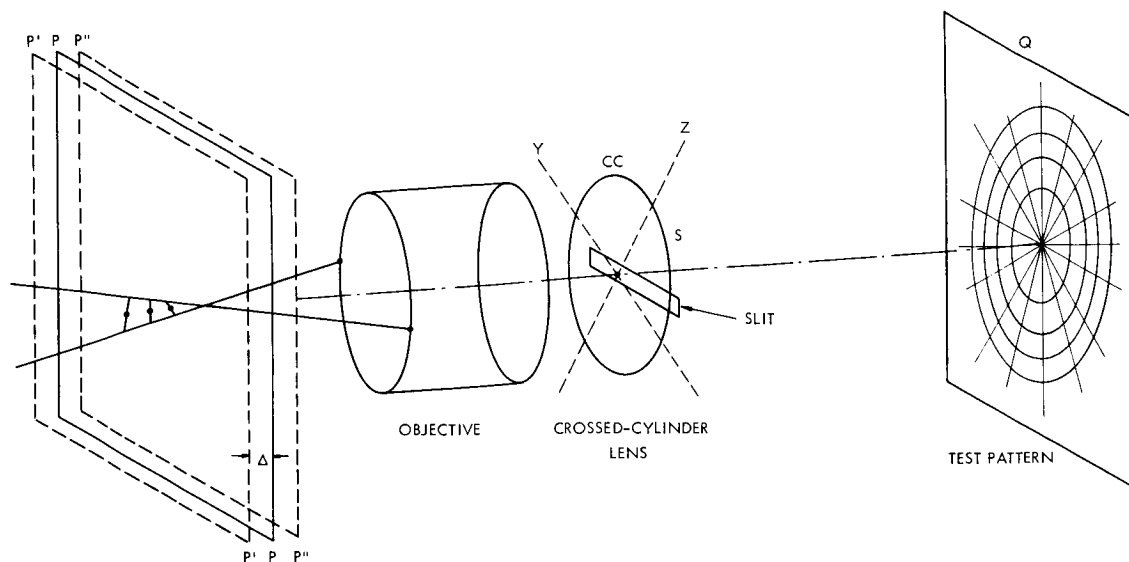
Fig. XIX-1. Arrangement of optical elements for testing the lens objective.

Accordingly, we have devised an alternative method of fabricating weak analyzer lenses, which makes use of the fact that the small section of the crossed cylinder that we actually use (that portion visible through the slit) can be approximated by a surface of a narrow rectangular glass prism stressed in torsion. A prism, 3 × 7 × ~100 mm is cut from the best quality of optical flat glass and stressed in torsion. Flexible couplings are used to eliminate bending moments. While the torsional stress is held constant, the prism is securely glued to a thicker optically flat glass window with a clear, well-mixed, epoxy cement (see Fig. XIX-2). The slit (parallel to the prism) is glued to the other side of this window.

Satisfactory long-term bonding can only be achieved if the glass surfaces are pretreated with a 2% solution of Union Carbide A-1100 Silane bonding agent because the glue joint must withstand indefinitely the torsional prestressing of the prism. Satisfactory bonds of 6-months endurance have been obtained by using R-314 Epoxy bonding agent (Carl H. Biggs Company) and also Epon 825 cement with a 20% Epon 2807 hardening agent. The maximum equivalent cylindrical power of the lens that we have fabricated in this way is ±1/8 diopter. We have found that analyzer lenses with 2-inch apertures and of successively halved powers of ±1/2, ±1/4, ±1/8, and ±1/16 diopters will suffice to test most camera lenses. We recommend the use of ophthalmic cylindrical lenses at the two strongest powers and the stressed-prism construction for the weaker powers.

We have examined the diffraction images of these new analyzer lenses when used with our best 200-mm telephoto lenses. Our tests with a high-power microscope
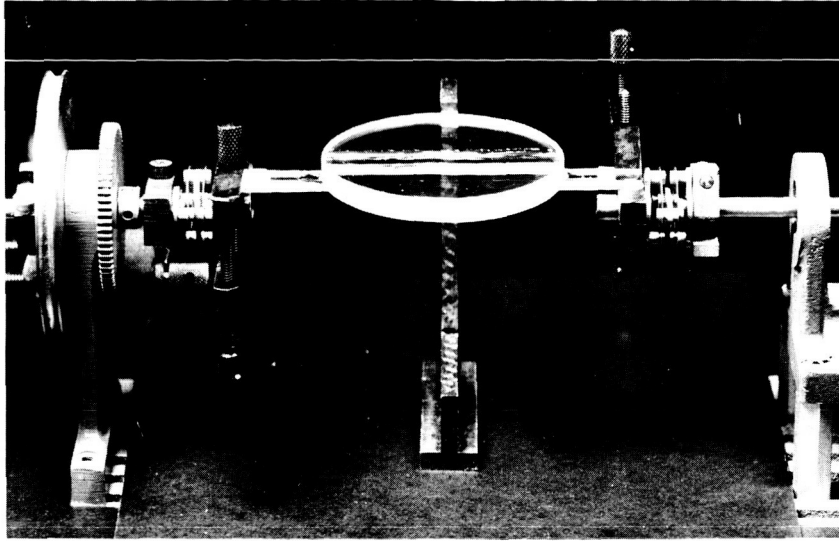
Fig. XIX-2.   Epoxy gluing of flat glass disc to prism of optical glass deformed by torsion. Flexible couplings minimize bending moments.

revealed some irregularities of performance which could be traced to the tele-photo lenses, but no defects that could be attributed to the new construction of the analyzer lens.

2.   Tests for Spherical Aberration and Coma

We are concerned with a lens test in which the image of a point source is a line segment. The angular orientation of this line-segment image is a sensitive function of the focus of the lens. In testing for spherical aberration, we introduce a compli-cation, in that the various zones of the lens may have different focal lengths. The image of a point source in this instance will not be a line segment, but with simple spherical aberration it will be a sigmoidal curve. In order to provide satisfactory imaging of the polar-coordinate chart in this new situation, it is preferable that the chart be ruled with white lines on a black background. With this modification, we tested a lens known to have large spherical aberration; the results are shown in Fig. XIX-3. (In Fig. XIX-4, we show a similar test with a well-corrected lens.) The effect of the spherical aberration is evident both in the appearance of the radial lines, and in the splitting of the images of the circles. The appearance and angular extent of the split images of the circles provides an indication of the spherical aberration of the lens (both in sign and magnitude); however, before precise meas-urements are attempted by this means, it will be necessary to standardize the film exposure and development conditions and to obtain correlation with other methods
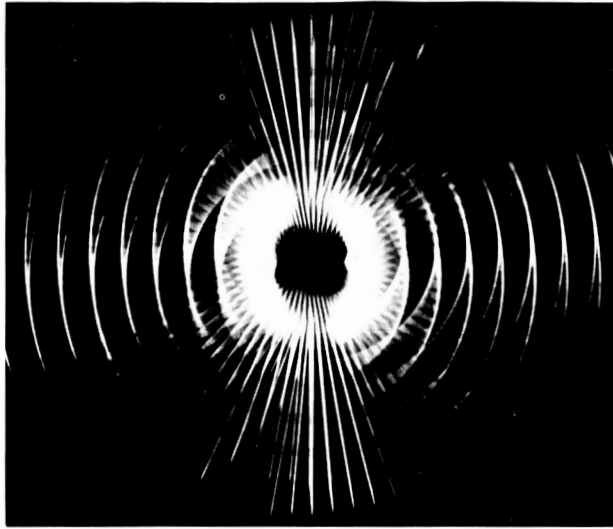
Fig. XIX-3.   Lens with appreciable spherical aberration.
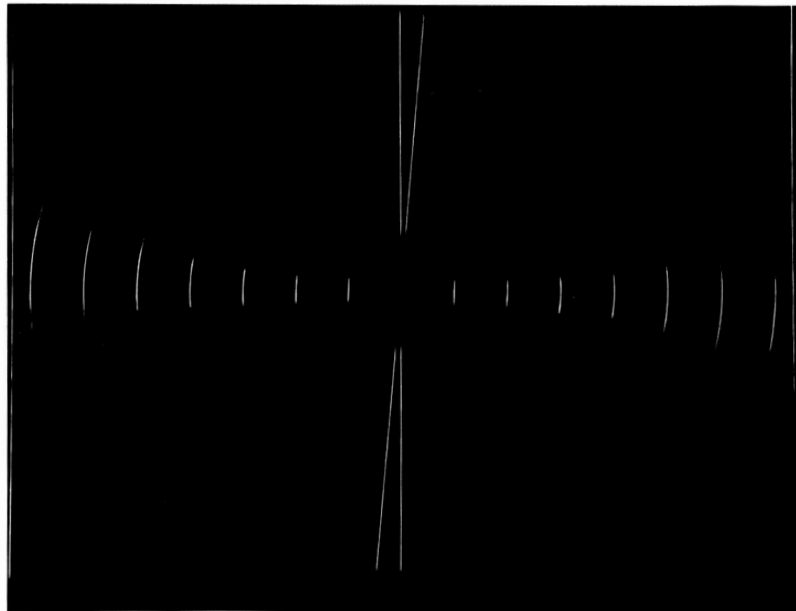(Compare with Fig. XIX-4.)
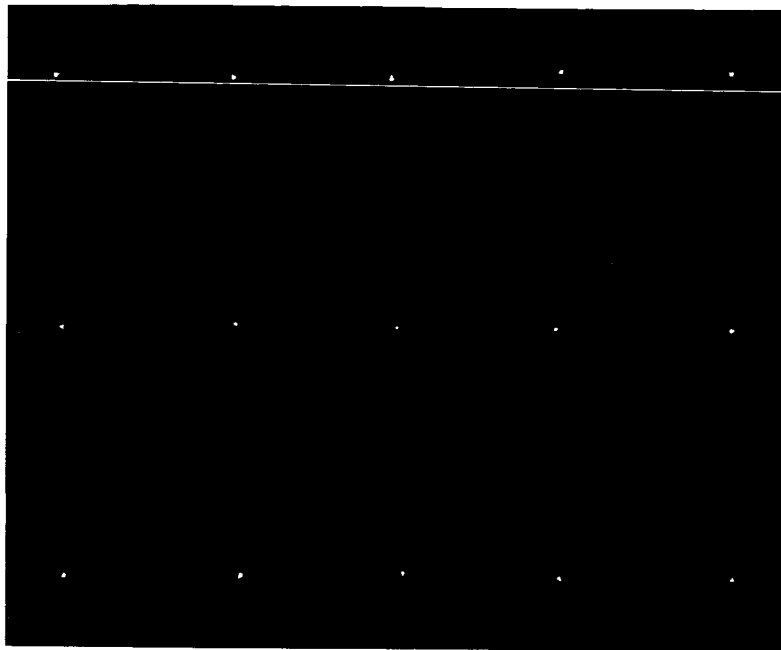

Fig. XIX-4.   Well-corrected lens.

Fig. XIX-5.   Array of point sources photographed
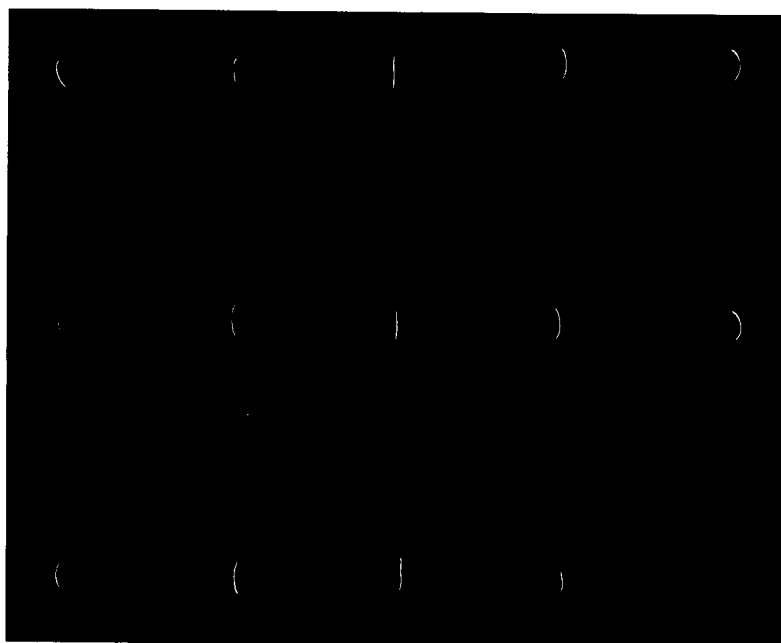with lens system having coma.

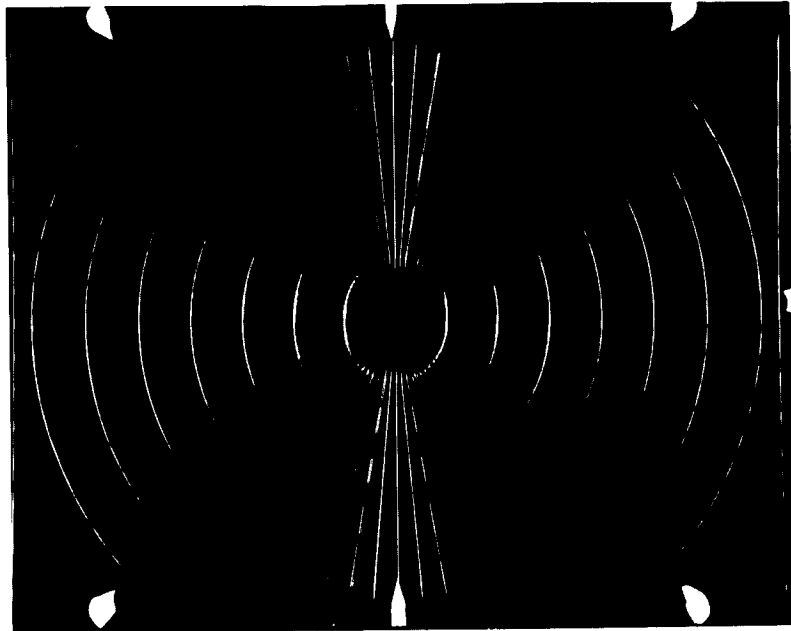Fig. XIX-6.   Use of lens analyzer to test for coma.

Fig. XIX-7. Test for coma with polar chart.
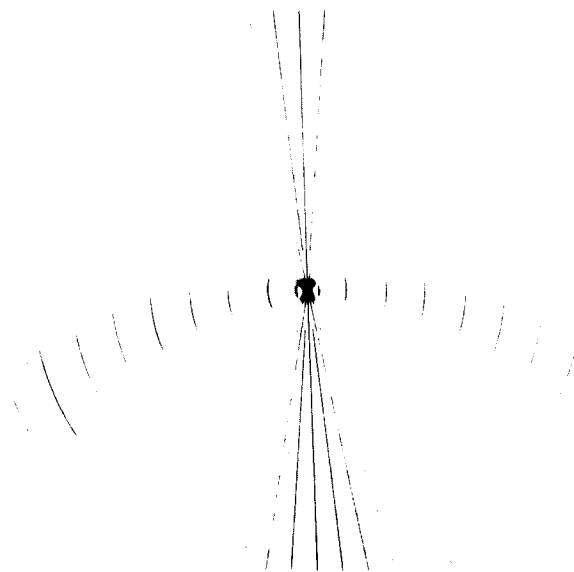(Compare with Fig. XIX-4).

Fig. XIX-8. Lateral chromatic aberration.

of measurement of the aberration of the lens.

The use of this method to determine coma is similar. We have simulated this aberration by the use of a TelXtender-type image amplifier, a Barlow-type negative amplifying lens interposed between the telephoto lens and the film plane. Such a combination has been found to generate almost pure coma, if the "TelXtender" is at full aperture. This is perhaps evident from the image of an array of 15 dots shown in Fig. XIX-5 (15 ball bearings illuminated with a single source). In Fig. XIX-6 we show the corresponding images photographed with the analyzer lens in place. This method is evidently a sensitive test for coma. In Fig. XIX-7 we show the appearance of the polar chart when photographed with the same apparatus. It is evident that the angular extent of the segments of the images of the circles becomes larger toward the edge of the field. Measurements of this angular broadening should give the magnitude — but not the sign — of the comatic distortion.

In summary, we note that this test determines four lens aberrations and also checks the focus and alignment of the camera. Critical focus is not necessary for this lens test to be effective.

3. Test for Lateral Chromatic Aberration

The method described here can be varied to permit determination of lateral chromatic aberration. The polar chart (black lines on white background) is photographed in white light with panchromatic high-contrast film through a weak dispersive analyzer prism oriented to provide spectral dispersion along the vertical axis of the chart. The exposure, development, and printing variables are so adjusted that only those line segments that point in the direction of the spectral dispersion will register on the prints. The spectral blurring in the vertical direction, caused by the analyzer prism, acts in combination with the radial spectral blurring of the image, which is the effect known as lateral chromatic aberration. The resultant of these two factors is a blur that obscures all but a selected set of the circumferential line segments. The figure formed by this set of line segments indicates the extent of the lateral chromatic aberration by the sign and magnitude of its curvature.

Since lateral chromatic aberration is not a common defect of modern lenses, we have simulated it by means of a large, single-element field lens placed one meter in front of our polar chart. This introduces a noticeable change in magnification with color, without significant focal error. For our analyzer prism, we used an ophthalmic prism that provides 10-cm deviation at one meter, and is made of glass with n = 1.523 and v = 58. In Fig. XIX-8 we show that the results of this test. The pattern has a similarity to our earlier test in which horizontal lens misalignment was indicated. It is important to note that in this test care must be taken with the focal adjustment.

B. Howland

B. QUANTUM THEORY OF MEASUREMENT IN RELATION TO A GENERAL THEORY OF OBSERVATION AND CONTROL*

1. Problem Statement

The quantum theory of measurement, due in its formal aspects to Dirac and von Neumann, is well known to have logical flaws, as well as to involve technical difficulties. Both originate from the nonrelativistic theory of stationary states; for it is immediately obvious that actual physical observables represent transitions between states with finite lifetimes which therefore cannot be stationary.

Logically, this shows up in von Neumann's account of measurement as the reduction of the triadic relation observer-instrument-object to a dyadic relation instrument-object. A standard example illustrating this point is the hydrogen atom; the electron, as object, is projected into a set of discrete states by the field of the proton as instrument. Nothing could actually be observed, however, without the perturbing effect of the radiation field (RF) which must be reckoned as the observer and form an integral part of the theory of measurement. The present perturbation theoretic treatment fails to be satisfactory because quantized fields such as RF cannot be defined unambiguously.

Apart from the theory of measurement, this last difficulty may be regarded as a technical one. Existing theory leans heavily on the use of self-adjoint operators which, with real eigenvalues, must represent stationary states. As a result, one practically has to adopt not only microcausality

$$[\phi(x), \phi(x')] = 0 \qquad (x-x')^2 < 0 \tag{1}$$

but also equal time commutation relations

$$[\phi(x), \pi(x')]_{t=t'} = \delta(x-x') \tag{2}$$

for field operators $\phi(x)$ and their conjugates $\pi(x)$. It is a well-known result[1] that there are uncountably many inequivalent representations satisfying (1) and (2). There then follows from the requirements of scattering theory the result known as Haag's theorem, namely the only unitarily equivalent representations are the free fields.

Our problem can now be stated as follows:

(A) Reformulate the theory of measurement in terms of triadic relations. In practice this means, find three term analogs of (1), (2).

(B) Show that (A) leads to a unique, or at least manageable equivalence classes of representation.

There is a distinct advantage in looking at this problem in a wider context than the physical, one can then ask what is unique about physics as a theory of observation; this will be implicit in the sequel. Before stating the approach specifically, some requisite background will be given.

2. Logic

There is a well-known isomorphism[2] between combinatory logic and computability, considered in terms of Turing machines. There is also the further reduction now known to exist (Arbib, Papert, unpublished lectures) to the theory of infinite, but finitely presented, semigroups. A basic building block of this theory is the Kleene three-term predicate $T(R, A, S)$ which states the existence of a logic (alternatively, a machine) with rules of inference R, axiom A, derivations S (in the machine's case corresponding indices, x, y, z, say). It is a well-known result that $\{x\} = (x \mid \exists_y T(x, x, y))$ is a recursively enumerable but nonrecursive set; that is, no finite amount of equipment will certainly decide whether $x_0 \in \{x\}$; nevertheless, $\{x\}$ is finitely generated and therefore has a well-defined structure. This is taken to indicate that in a theory of observation we may reasonably speak of nonobservable entities having a structural significance.

The discussion above may be exemplified in the logic of quantum mechanics represented in the conventional way[3] as the lattice of all closed subspaces of Hilbert space. This is known to be a relatively orthocomplemented lattice which is not modular,[4] that is, it has sublattices of the form shown in Fig. XIX-9.
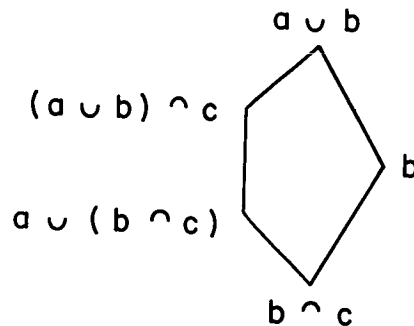

$$a \cup b$$
$$(a \cup b) \cap c$$
$$a \cup (b \cap c)$$
$$b$$
$$b \cap c$$

Fig. XIX-9.

Here

$$(a \cup b) \cap c \neq a \cup (b \cap c); \tag{3}$$

however, the essential features of existing quantum mechanics depend only on modularity (i. e., equality in (3)) without distributivity

$$a \cup (b \cap c) \neq (a \cup b) \cap (a \cup c) \tag{4}$$

$$a \cup b \cup b' \_\_\_$$

$$b \quad b' \quad b'' \_\_\_$$
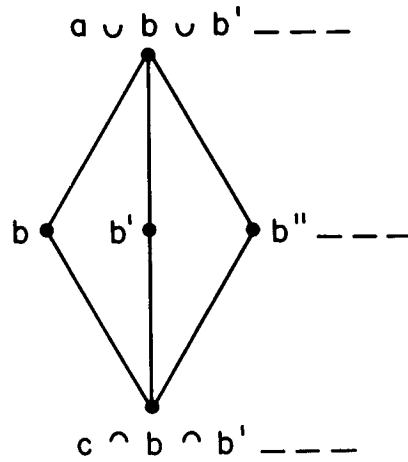
$$c \cap b \cap b' \_\_\_$$

Fig. XIX-10.

as in Fig. XIX-10. It seems very likely then that the ambiguity of field theory arises from insufficient restriction on the appearance of nonmodular sublattices, thereby indicating ordered indiscernibles rather than equivalent indiscernibles (as in the modular case, Fig. XIX-10), and this restriction represents some symmetry condition on the predicate T(R, A, S).

3. Quantum Field Theory

The typical problem[5] of field theory is the solution of the nonlinear equation

$$(\Box + k^2) \Phi = \epsilon^2 \Phi^3 \tag{5}$$

where Φ is an operator-valued function subject to conditions (1) and (2). The existing approach is to treat the right-hand side as a perturbation and expand the integral form of solution

$$\Phi(x) = \Phi_0(x) + \epsilon^2 \int \Delta_0(x-x') \, \Phi(x') \, dx', \tag{6}$$

where Φ_0, Δ_0 are solutions of the 'free' equation ($\epsilon=0$). To remove divergences from the expansion, it is necessary to interpret it in terms of Feynman diagrams; while this can be done consistently, it is ad hoc and leaves us with no theroy about what determines the experimental values of (m, g). There are other equal methods such as the program known as 'Reggeization' in scattering theory of removing ambiguities from field theory.

It seems that there are at least two other approaches that could lead to a deeper theory. The first is to consider (5) as an equation in the non self-adjoint operators

$$A^{\pm} = (J \pm ik \pm \epsilon \, \Phi). \tag{7}$$

To justify this one would have to show that the A's give a representation of a group that expresses, in some basic way, a theory of measurement. This will be indicated in section 4.

Alternatively, one can notice that if ϕ is a c-number, (5) has a solution in terms of elliptic functions with modulus τ (period ratio) which determines $(m(\tau), g(\tau))$. The last need not be constant off the light cone (or mass shell) and the modular transformation group Γ gives a basis for introducing what is currently known as the renormalization group.[6] To go over to a g-number theory one can consider the Hilbert space of differentials on a Riemann surface[7] and its various orthogonal decompositions. The importance of this approach is that it establishes a connection with the powerful methods of algebraic field theory when $K(\tau)$ is algebraic. In particular, we shall have in some limit

$$\omega g \left| \tau \right|_v = (x-x')^2, \tag{8}$$

where v is a valuation that may not be Archimedean.[8] This crucial point explains why nonlocal field theories have failed up till now. A typical[9] form of it requires relations of form

$$[P^\mu [P_\mu, \psi]] = k^2 \psi$$

$$[X^\mu [X_\mu, \psi]] = \lambda^2 \psi; \tag{9}$$

however, if X is chosen to be a self-adjoint (or even normal) operator (therefore diagonalizable, with eigenvalues x^μ, say) we are back essentially to a localizable theory, since there is no way of distinguishing the domains $x^\mu x_\mu \gtrless \lambda^2$.

4. Non Self-Adjoint Operators and Group Representations

If A is not a normal operator (i. e., does not commute with its adjoint) then it cannot be unitarily equivalent to a diagonal, but only a triangular form. The eigenfunctions are not orthogonal and this has the apparently unpleasant consequence that probability is not conserved, but there is no evidence that this should be so in every bounded domain. Non self-adjoint operators have forced themselves on the attention of physicists,[10] in particular, in the theory of scattering by metastable systems. Here we should ask only that the A's give a representation, necessarily nonunitary, of some group that is uniquely associated with a theory of measurement. Mackey[11] has pointed out that the most natural infinite representations of groups are not necessarily unitary and that nonunitary representations can be induced for subgroups by unitary representations of larger ones. Two questions now arise which can be seen as reformulations of (B) and (A), respectively.

(C) What is the source of nonunitary representations? The answer is in algebraic groups,[12] that is, groups whose field of definition is restricted to algebraic numbers. This is discussed further in section 5.

(D) What group structure expresses the uniqueness of physical measurement? The answer here, which I have discussed in a preliminary fashion elsewhere,[13] lies in a

phenomemon occurring uniquely in the exceptional simple Lie groups which is known as the 'Principle of Triality'.[14,15]

This principle can be formulated as a physical one by regarding it as a far-reaching extension of the 'Dirac trick,' by which is meant the following — in the relativistic theory of the electron, the angular momentum operators M (as generators of the Lorentz group D_2) are extended to M' = (M+σ), so that M' becomes a constant of the motion. In this case M' acts on a direct product of two representations (Infinite and Spin) of the same group D_2. We have a degeneracy which would be lifted if the direct product representation were replaced by a single representation of a larger group; it seems that there is a unique candidate for the latter, namely the exceptional Lie group F_4. Operators of this group can be put in the form A = $(Z+\sigma_1+\sigma_2)$, where the terms operate respectively on a vector space V and two spin spaces S_1, S_2. There exists, then, as an automorphism of the group, a triality operator J such that V = $JS_1 = J^2S_2$. The existence of J is due to the fact that V is an 8-dimensional vector space (relativistic phase space) that can be coordinatized by the nonassociative division algebra $Q_8(z)$ of octonions which, in addition to the involution Tz = $\bar{z}$, has another $S(z_1z_2) = z_1(z_2\cdot)$ because of the nonassociativity. Then J = ST.

Triality is the mathematical expression of what has been called[16] 'the elementarity of measurement,' since it says — no matter how V is distinguished, the information obtained is the same. The group F_4 thus contains within it not only coordinate but gauge transformations that at present are considered separately, for example,[17] as the group A_2 of strong interaction symmetries.

It follows from this that, unlike the Dirac case, the representations of σ_1, σ_2 must be infinite dimensional, at least if that of Z is. But, in any case, there are no finite dimensional representations of $Q_8(z)$ on account of the nonassociativity. The nature of these representations is at present a matter of speculation, but in view of what was said in the previous section it seems likely that they may be totally discontinuous p-adic representations. If this is the case, possible substitutes for (1) would be

$$[\sigma_1, \sigma_2] = Jv(z-z'), \qquad (10)$$

together with two further relations obtained by application of J, since this can be assumed to act on the valuation of v(z).

5. Algebraic Field Theory

The key to representation theory is algebraic groups.[20] There is a classical approach to these through Abelian functions that represent an algebraic field on a Riemann surface. A surface of genus g (i.e., g cuts) is completely characterized by (3g-3) moduli. The transformation group Γ on these moduli establishes a correspondence[19] between surfaces which, when the moduli are from an algebraic

number field $K(\tau)$, may be an automorphism (for elliptic functions this is known as complex multiplication). To represent a Lie group G on the variables of these functions one has to embed the Weyl group $W(G)$ in Γ. Elsewhere[18] I have given some conjectures about this for F_4. An essential role here is played by the class field theory of $K(\tau)$, that is, the equivalence classes of ideals in $K(\tau)$, hence the connection with p-adic valuations. Since Γ is an infinite group, it may contain finitely presented groups with an unsolvable word problem, hence a connection also with the remarks on logic of section 2.

It seems quite essential to preserve the preferred role of time as implied by (2), and therefore another condition imposed on $K(\tau)$ will be that it causes just such a splitting of the group structure.

6. Global Viewpoint

The group F_4 represents only half, which we may consider the local part, of the picture. If we write the relativistic wave equation in the form

$$\left\{\Gamma^\mu \left(\nabla_\mu + i \frac{\delta}{\delta J^\mu}\right) + i\kappa\right\} \psi = 0, \tag{11}$$

where the covariant derivative

$$\nabla_\mu = \partial_\mu + i\langle A_\mu\rangle \tag{12}$$

is determined by the expectation value $\langle A_\mu\rangle$ of the electromagnetic potential, then the triality$\left(\frac{\delta}{\delta J^\mu}, \nabla_\mu, \kappa\right)$ is local in the sense that it describes events in the hypersurface determined, apart from $\langle A_\mu\rangle$, by the element of support Γ^μ considered as a constant. If the last is no longer assumed, we have the point of view of general relativity and all the arguments about triality now apply to Γ^μ.

F_4 is not the largest group that acts on an octonion structure. This structure is E_8 which acts on a tensor product of octonion algebras in the same way that D_6 acts on the Dirac algebra (of Γ's) as a tensor product of quaternions. It is evident that the splitting relations here are important just as they are in the Dirac case, where

$$(\Gamma^\mu, \kappa) = (\sigma_1, \rho_i, 1, \rho_3) \tag{13}$$

(σ_i, ρ_i, elements of complex quaternion algebras). In particular, it is clear, in view of (12), that the 'trialities' will not be independent.

The relation between the groups is[15]

$$m = \quad 1 \quad 2 \quad 3 \quad 8$$

$$n = 1 \quad B_1 \quad A_2 \quad C_3 \quad F_4$$

$$2 \quad A_2 \quad A_2 \quad A_5 \quad E_6$$

$$4 \quad\quad\quad\quad D_6 \quad E_7$$

$$8 \quad\quad\quad\quad\quad E_8$$

Here, (m, n) are the dimensions of the algebras; B_1 is the rotation group of real 3-space; A_2, A_5 (=SU(3), SU(6)) have been proposed as gauge groups for strong interactions. If A_5 turns out to be the correct identification, then we have the possibility of the approximate splitting, $E_8 \sim A_5 + B_3$, where B_3 is the smallest group admitting a nontrivial triality, for which I have given an interpretation elsewhere in terms of a cosmological model.[22]

M. C. Goodall

References

1. G. Barton, Advanced Field Theory (Interscience Publishers, Inc., New York, 1963).

2. M. Davis, Computability and Unsolvability (McGraw-Hill Book Company, New York, 1958).

3. G. W. Mackey, Mathematical Foundation of Quantum Mechanics (W. A. Benjamin, Inc., New York, 1963).

4. Garrett Birkhoff, Lattice Theory, Proc. Symp. Pure Math. $\underline{2}$, 156 (1961).

5. A. S. Wightman, Relativistic Quantum Field Theory, Proc. International Congress on Mathematics, 1962.

6. N. N. Bogoliubov and D. V. Shirkov, Theory of Quantized Fields (Interscience Publishers, Inc., New York, 1959).

7. L. Ahlfors and L. Sario, Riemann Surfaces (Princeton University Press, Princeton, N. J., 1960).

8. G. Bachman, p-adic Numbers and Valuation Theory (Academic Press, New York, 1964).

9. H. Yukawa, Non-localisable field theory, Phys. Rev. $\underline{77}$, 219 (1950).

10. C. L. Dolph, Non-selfadjoint problems in mathematical physics, Bull. Am. Math. Soc. $\underline{66}$, 1 (1960).

11. G. W. Mackey, Infinite dimensional group representations, Bull. Am. Math. Soc. $\underline{69}$, 628 (1963).

12. I. Barsotti, Algebraic group varieties (Bull. Am. Math. Soc.)

13. M. C. Goodall, Foundations of relativistic quantum mechanics, Nature $\underline{195}$, 167 (1962).

14. C. Chevalley, Algebraic Theory of Spinors (Columbia University Press, New York, 1954).

15. H. Freudenthal, Lie groups in the foundations of geometry, Adv. in Math. $\underline{1}$, 2 (1964).

16.   M. Sachs, Elementarity of Measurement in General Relativity, Boston University Colloquium paper, 1965.

17.   S. L. Glashow and M. Gell-Mann, Gauge theories of vector particles, Ann. Phys. 15, 437 (1961).

18.   M. C. Goodall, Representation of a Group in Relativistic Quantum Mechanics, Nature 197, 585 (1963).

19.   F. Conforto, Abelsche Funktionen u. Algebraische Geometrie (Springer Verlag, Berlin, 1956).

20.   A. Weil, Adeles and Algebraic Groups (Lecture Notes, Institute for Advanced Study, Princeton, N. J. (1961).

21.   M. Jacob and G. Chew, Strong Interaction Physics (W. A. Benjamin, Inc., New York, 1964).

22.   M. C. Goodall, Connexion between local and global physics, Nature 197, 994 (1963).

L. Stark
J. F. Dickson III
G. Cook
D. Martin
G. A. Masek

J. A. Michael
H. Miyahara
A. Natapoff
J. Negrete
O. Sanchez-Felipe
A. A. Sandberg

S. F. Stanten
T. Webb
G. H. Whipple
G. N. Yankelevich
B. L. Zuber

A. MICROSACCADES AND THE VELOCITY-AMPLITUDE RELATIONSHIP FOR SACCADIC EYE MOVEMENTS

It has been known for some time that the velocity of a saccadic eye movement is dependent upon its amplitude.[1-3] This relationship is one of the nonlinearities encountered in a study of the versional eye movement system as a servomechanism.[4] Westheimer[1] has studied the relationship of maximum velocity to amplitude for voluntary saccades ranging from 2 to 30 degrees. His curve shows a curvilinear relationship over the entire range, with a tendency to saturation for larger amplitudes. Gurevich,[2] in studying average velocity, has found a similar relationship, although his velocities are much lower than Westheimer's, as would be expected. Gurevich also found that average velocity measurements for any given size movement were fairly constant under the following conditions: horizontal, diagonal or vertical movement; variation of starting position and direction of movement; in the presence of visible fixation points for the stimulus or in total darkness with conditioned eye movements. He also found that average velocities of secondary saccadic corrections fell on the same curve that he obtained for the types of movements described above. The range of amplitudes used in his study was from 1 to 35 degrees.

The data of Gurevich indicate that a single physiological system is responsible for a wide variety of saccadic eye movements. In an attempt to determine if microsaccades, the small (1 to 20 minutes of arc) involuntary saccades observed during fixation, are the output of this same system, we have made a study of the maximum velocity of such movements.

The apparatus for presenting a fixation point and calibration points at optical infinity, as well as the method of measuring these small eye movements, has been described[5] previously. The signal proportional to eye position was recorded on one channel of a recorder (Sanborn Model 320). This signal was also electronically differentiated and the derivative recorded on the other channel of the recorder. In an attempt to improve resolution, signals proportional to the pen positions on the eye position and velocity channels were recorded on a second recorder (Visicorder Model 1508). This provides a

further amplification factor of approximately five. Thus two recorders were used, the Sanborn being primarily used to keep both signals on scale and to provide an immediate check on the linearity of calibrations. Visicorder records were used in all analyses.

Calibration of the velocity channel was accomplished by recording a triangular wave on the eye-position channel and its derivative on the velocity channel. All gains and calibrations were unchanged for this procedure. Thus, given the amplitude of the triangular wave on the eye-position channel and the frequency of the waveform, a velocity in degrees per second could be related to a given deflection on the velocity channel. Such
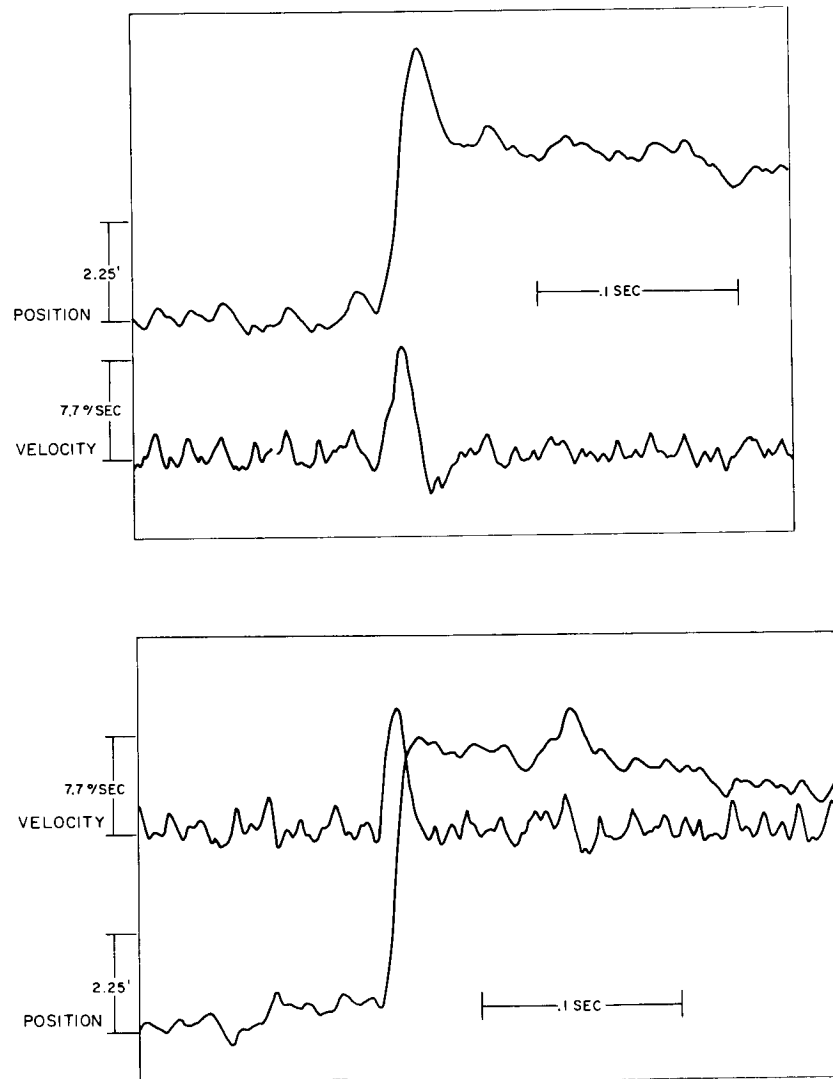
Fig. XX-1. Two typical microsaccades and their velocity traces. Note that although the two movements are approximately the same size, only one has overshoot.
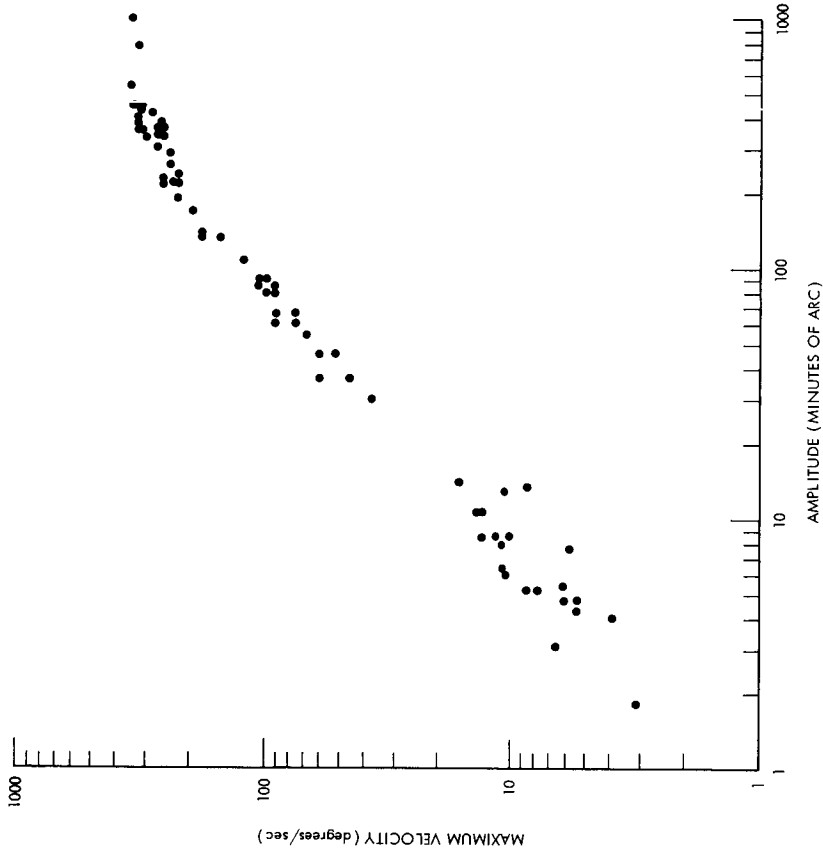
Fig. XX-3. Maximum velocity (degrees/second) vs amplitude (minutes of arc) for microsaccades, involuntary corrective saccades, and voluntary saccades.
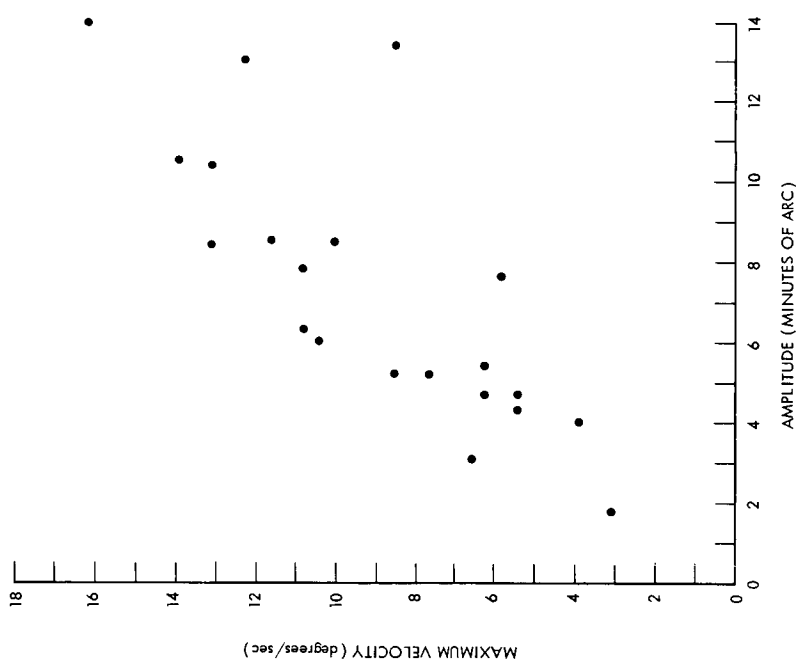
Fig. XX-2. Maximum velocity (degrees/second) vs amplitude (minutes of arc) for microsaccades.

calibrations were made for at least three frequencies within the range of velocities observed in the experiment.

Figure XX-1 shows some typical microsaccades and their velocity traces. Here we see two movements of roughly the same size, one with a great deal more overshoot than the other. Note that the overshoot is proportionately much greater than that normally seen with larger saccades. Figure XX-2 is a plot of maximum velocity in degrees per second (ordinate) as a function of amplitude in minutes of arc (abscissa). It is clear that velocity is an increasing function of amplitude for these movements.

In Fig. XX-3 we have replotted the data from Fig. XX-2 and added data points from larger voluntary saccades and secondary corrective saccades. The latter data were obtained in the same manner as those for the microsaccades, except, of course, that the stimulus conditions were different. The velocity data for the larger movements are very close to those of Westheimer.[1] The points are plotted on logarithmic scales because of the large ranges involved. A smooth continuous curve through all data points is clearly justified, and indicates, indeed, that microsaccades are produced by the same physiological system as voluntary saccades and involuntary corrective saccades.

It is interesting that so much overshoot is observed on microsaccades as compared with larger saccades. With further experiments designed to provide a dynamical model of this system, it is hoped that some explanation of this observation will be forthcoming.

We are indebted to Professor L. R. Young, of the Department of Aeronautics and Astronautics, M. I. T., who pointed out the need for velocity data on microsaccades.

B. L. Zuber, G. Cook, L. Stark

References

1. G. Westheimer, Mechanism of saccadic eye movements, A. M. A. Arch. Ophthalmol. 52, 710 (1954).

2. B. Kh. Gurevich, Universal characteristics of fixation eye jerks, Biofiz. 6, 377 (1961).

3. J. E. Hyde, Some characteristics of voluntary human ocular movements in the horizontal plane, Am. J. Ophthalmol. 48, 85 (1959).

4. L. R. Young and L. Stark, A discrete model for eye tracking movements, IEEE Trans. Vol. Mil-7, p. 113, 1963.

5. B. L. Zuber, A. Crider, and L. Stark, Saccadic suppression associated with microsaccades, Quarterly Progress Report No. 74, Research Laboratory of Electronics; M. I. T., July 15, 1964, p. 244.

B. HUMAN HORIZONTAL EYE-MOVEMENT MECHANISM

This report gives an account of a continuing investigation into the horizontal eye-movement mechanism, the first part of which was reported in Quarterly Progress

Report No. 76 (pages 343-352). Since the time of the first report, new evidence has brought about certain changes and additions in the model of the physical plant.

The updated model (Fig. XX-5) has been used with experimental data to determine the controller or driving-function behavior during saccadic movements. The minimum time behavior for the model has also been calculated and is compared with the actual behavior.

1. Physical Structure

a. Resting Length of Muscle

Robinson[1] describes measurements on the lateral rectus of a cat in which he finds the rest length to be approximately 25 mm. By extrapolation from a figure of Robinson,[2] one finds that the distance from the length of maximum tension to that of zero tension is approximately 12 mm. From Wilkie,[3] we find that this distance just described is $L_o/2$, where L_o is the length of the muscle for developing maximum tension. Therefore, $L_o/2 \approx 12$ mm, and $L_o = 24$ mm. So the rest length is approximately the same as L_o which means that if we look straight ahead, the operating point of the muscle is on the peak of the curve. See Fig. XX-4.
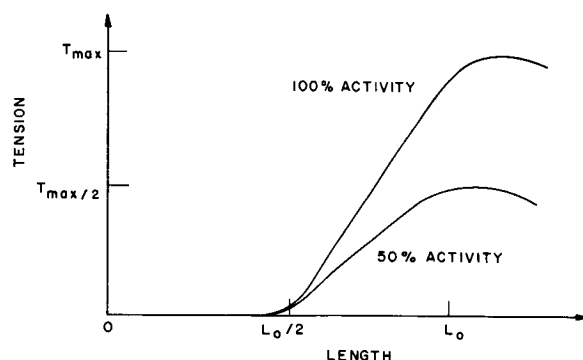
Fig. XX-4. Active muscle length-tension relationship.

This point brings up a question about stability, since the stretched muscle will now be operating with a negative spring coefficient. The shortened muscle, which is operating with a positive spring coefficient, is more highly activated and predominates, thereby ensuring stability at any resting position. As for the dynamic situation, during a movement toward center where the stretched muscle is more highly activated than the shortened one, incremental instability may exist. This means that for a given degree of activity the force pulling on the eye increases during the movement, although the eye is moving in the direction of the force. As the desired position is approached and the

activity of the stretched muscle falls off, the stable situation described above prevails.

A factor that lessens this temporary incremental stability is the flatness of the length-tension muscle characteristic in the neighborhood of the peak. Because the
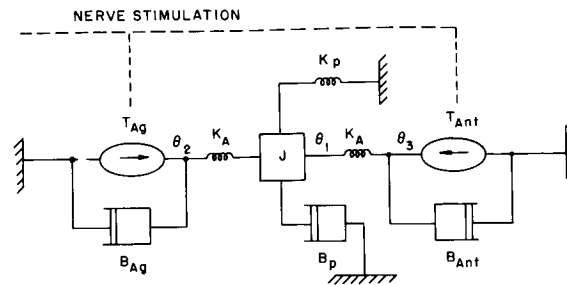

Fig. XX-5. Model of the physical system.

muscle does operate in this flat region, the contractile element is modeled as a tension source in parallel with nonlinear damping (Fig. XX-5).

b. Passive Tension

With the active muscles represented as tension sources with appropriate dynamics, the passive tension becomes important in bringing about an equilibrium of forces when the two muscles have different degrees of activity. From Ruch and Fulton,[4] we find the length tension slope of a relaxed muscle to be T_{max}/L_o. For the human, assuming $T_{max} \approx 500$ gm and $L_o \approx 4$ cm, we obtain $Kp \approx 125$ gm/cm; and since the radius of the eye is approximately 1 cm, $Kp \approx 2$ gm/degree.

c. Active Muscle Elasticity

Wilkie[3] states that at $T_o = T_{max}$, 3 per cent stretch occurs in the active muscle. For $T_{max} = 500$ grams, and a length of 4 cm, this yields a spring coefficient of

$$K_A = \frac{500 \text{ gram}}{0.03 \times 4 \text{ cm}},$$

or since 1 cm $\approx$ 1 rad, $K_A = 73$ grams/degree.

d. Active Damping

Hill,[5] who has spent many years studying muscle behavior, advanced the most accepted analytic description of a shortening active muscle behavior in the following equation:

$$v(T+a) = b(T_o - T),$$

where T is tension, v is velocity, T_o is static tension, and b is one-fourth the maximum shortening velocity of the muscle. This equation can be rearranged to yield

$$T = T_o - \left(\frac{T_o + a}{v + b}\right) v.$$

Now, $a \approx 0.25\, T_o$. Therefore

$$T = T_o - \left(\frac{1.25\, T_o}{v + b}\right) v,$$

and the active damping coefficient is

$$B_{Ag} = \frac{1.25\, T_o}{v + b}.$$

Here, T_o is the tension with zero velocity and with the muscle at length L_o. T_o increases with increased activity to a maximum value of T_{max}. We then see that the active damping coefficient of a shortening muscle is proportional to the degree of activity and is also a function of velocity.

It can be shown from the model, Fig. XX-5, that during an isometric contraction,

$$b = 0.25 \frac{\dot{T}}{K_A}.$$

Now,

$$K_A = \frac{T_{max}}{0.3\, L},$$

where L is the rest length of the muscle. Therefore

$$b = 0.0075 \times \frac{\dot{T}}{T_{max}} \times L.$$

Assuming similar time behavior between cats and humans, we refer to Adler.[6] He shows an isometric contraction of a cat; $L \approx 2.5$ cm, $\dot{T}/T_{max} = 210$, and $b = 3$ rad/sec or $180°$/sec. Recalling that $4b$ = maximum muscle velocity, we will take $b = 250°$/sec, since velocities as high as $800°$/sec have been reported.

Katz[7] showed that the damping coefficient in an active muscle that is being lengthened is quite different from that in one that is shortening. It is still proportional to activity, but much larger than that for the shortening muscle and no longer a function of velocity. From Katz[7] we find

$$B_{Ant} = \frac{12.5}{b}\, T_o.$$

Furthermore, Katz states that forces greater than $1.8\ T_{max}$ in the lengthening muscle can damage the muscle. This factor will be an important consideration in calculating the minimum time behavior of the model.

e. Primary Position Tension

Breinin[8] shows a graph of integrated muscle potential vs eye position. Linearizing this curve, we obtain Fig. XX-6. For a 15° movement, the difference in tension

$$T_{Ag} - T_{Ant} = 15 \times Kp = 30 \text{ grams.}$$

In Fig. XX-6 this corresponds to 6.4-4.8, or 1.6 units on the ordinate. Inman et al.[9] found a linear relation between electric integral and tension with constant muscle length.
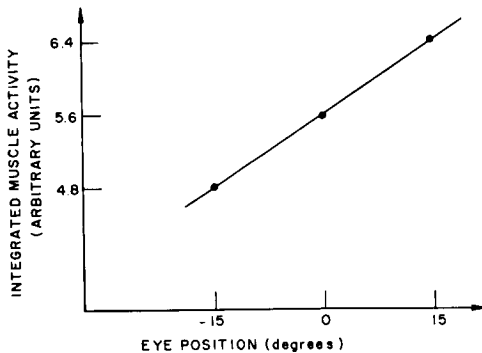
Fig. XX-6. Integrated muscle activity vs eye position.

We shall assume that the small length changes, 15°, do not invalidate this relation which we shall now utilize.

$$\frac{30 \text{ gm}}{1.6 \text{ units}} \times 5.6 \text{ units} = 105 \text{ grams.}$$

Thus, the tension in each muscle in the primary position, 0°, is approximately 100 grams.

f. Passive Damping

An experiment by Robinson[1] was used to determine the passive damping. The eye was activated for the zero-degree position, forced to the 4° position and released. Upon release, if we neglect inertia, it can be shown that $Bp\dot{\theta}(O) = Kp\theta(O)$.

With $Kp = 2$, $\theta(O) = 4$ and $\dot{\theta}(O) = 520°/\text{sec}$, we obtain $Bp = 0.016$ gram sec/degree.

For convenience, all the parameters of the system illustrated by Fig. XX-5 are tabulated:

$$B_{Ag} = T_{Ag}\left(\frac{1.25}{250 + \dot{\theta}_2}\right)\frac{\text{gram sec}}{\text{degree}}$$

$$B_{Ant} = T_{Ant}\frac{12.5}{250}\frac{\text{gram sec}}{\text{degree}}$$

$$K_A = 73 \text{ grams/degree}$$

$$Kp = 2 \text{ grams/degree}$$

$$Bp = 0.016 \text{ gram sec/degree}$$

$$J = 0.000047 \text{ gram sec}^2/\text{degree}.$$

2. Controller Behavior

The basic configuration is shown in Fig. XX-7. The target and output are known, and we would like to know the behavior of the control variable, U

$$W(U) = \theta_{out}$$

$$U = W^{-1}(\theta_{out}).$$

If W is known and if W^{-1} exists, then U can be found. Here, U is the nerve stimulation to the agonist and antagonist muscles.

This can be measured directly in the form of electromyograms. The accuracy and
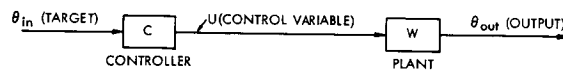
Fig. XX-7. System configuration.
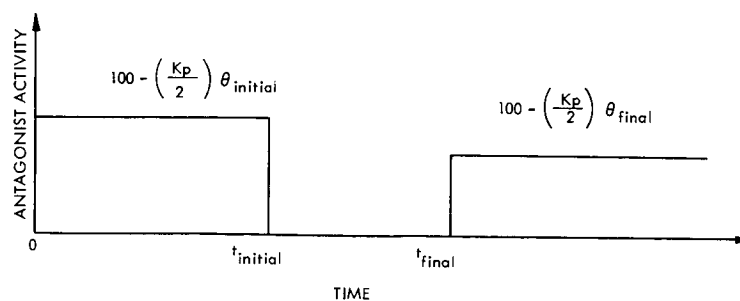
Fig. XX-8. Antagonist behavior during a saccadic movement.

readability of such measurements, however, leaves much to be desired, except in static situations.

Since W is known and W^{-1} exists, experimental output data can be used to calculate U. A problem does exist in that there are two unknowns, T_{Ag} and T_{Ant}. The electromyograms of the antagonist (see Davson[10]) are taken as the more readable of the two. On this basis, T_{Ant} is assumed to have the behavior shown in Fig. XX-8. This leaves T_{Ag} as the only unknown and it can be determined.

In the form assumed for T_{Ant}, we require that before the movement

$$T_{Ant} = 100 - \left(\frac{Kp}{2}\right)(\theta_{initial});$$

and after the movement,

$$T_{Ant} = 100 - \left(\frac{Kp}{2}\right)(\theta_{final}).$$

We also desire that during the movement, $T_{Ant} \approx O$.

This was programmed on the computer first by generating a function g_1 such as that shown in Fig. XX-9 with the corners coming right after the start of the movement and right before the end of the movement. This function g_1 was then divided by $1 + K\dot{\theta}$ to yield g_2; that is, $g_2 = \dfrac{g_1}{1 + K\dot{\theta}}$. K was taken to be of such size as to cause g_2 to be very small throughout the movement.
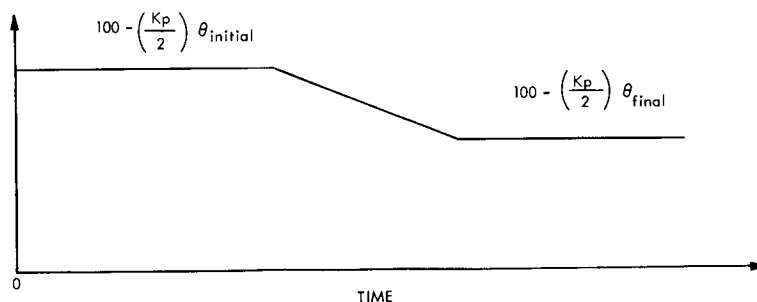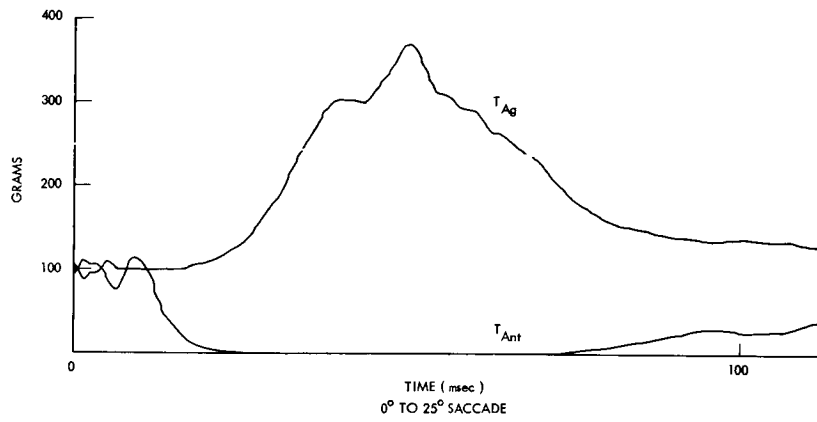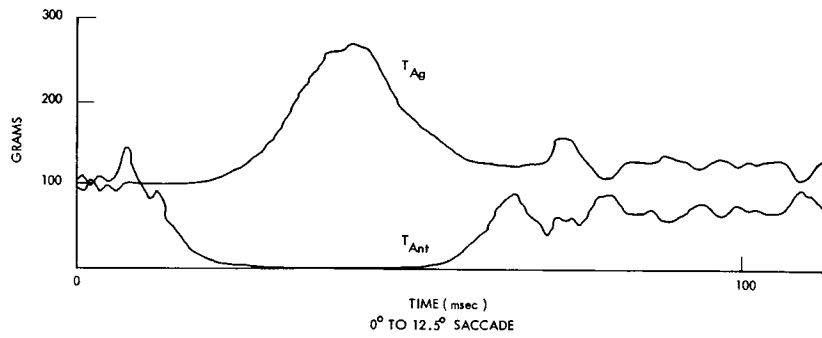

Fig. XX-9. g_1 artificial function used in generating T_{Ant}.

The behavior of T_{Ant} is shown in Fig. XX-10, together with the resulting behavior of T_{Ag}. It is seen that sometimes there is a burst of activity in T_{Ant} toward the end of the movement. This is caused by overshoot in the movement ($\dot{\theta}$ going negative) and

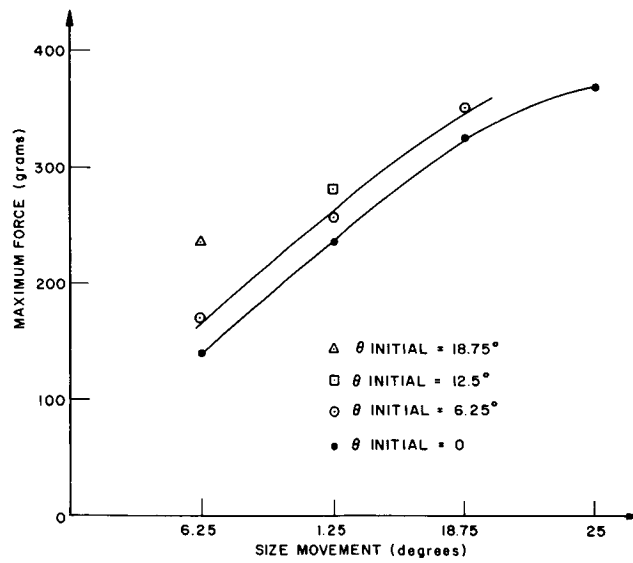Fig. XX-10. Computed controller behavior during
saccadic movements.

Fig. XX-11. Computed maximum force during
saccadic movements.

by the way g_2 is generated, that is,

$$g_2 = \frac{g_1}{1 + K\theta}.$$

Davson[10] reports that such bursts have been observed in the antagonist, and, for this reason, no attempt was made to remove the peak.

The reciprocal innervation is well displayed here. The maximum force applied by the agonist for various movements is tabulated below and displayed in Fig. XX-11.

Movement (degrees)	Max T_{Ag} (grams)
0 to 6.25	143
0 to 12.5	240
0 to 18.75	323
0 to 25.0	369
6.25 to 12.5	172
6.25 to 18.75	254
6.25 to 25.0	337
12.5 to 25.0	283
18.75 to 25.0	243

3. Minimum Time Behavior

The control function that would be applied to the system to attain a given state in minimum time has been determined. Position and velocity plots, together with a phase trajectory for such a movement, are shown in Fig. XX-12. These data were obtained through computer simulation. The control behavior during these movements is as follows:

$$T_{Ag}[0, t_1) = 100 + \frac{Kp}{2}\, \theta_{initial}$$

$$T_{Ag}[t_1, t_s) = 500 \text{ grams } (T_{Ag}\max)$$

$$T_{Ag}[t_s, t_2) = 0$$

$$T_{Ag}[t_2, \infty) = 100 = \left(\frac{Kp}{2}\right)\theta_{final}$$

$$T_{Ant}[0, t_1) = 100 - \left(\frac{Kp}{2}\right)\theta_{initial}$$

$$T_{Ant}[t_1, t_s) = O$$

$$T_{Ant}[t_s, t_2) = 500 \text{ grams } (T_{Ant} \text{max})$$

if $\dot{\theta} \leq 16°/\text{sec}$

$$T_{Ant}[t_s, t_2) = \frac{720}{1 + \frac{\dot{\theta}}{20}}$$

if $\dot{\theta} > 16°/\text{sec}$

(Recall $T_{Ant}\left(1 + \frac{\dot{\theta}}{20}\right) \leq 720$ to prevent muscle damage.)

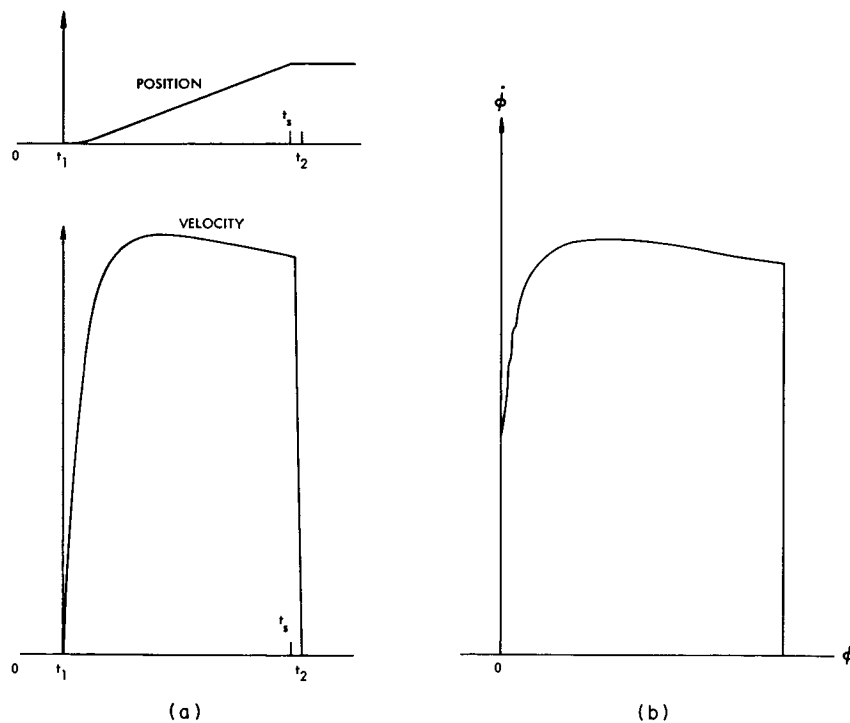$$T_{Ant}[t_2, \infty] = 100 - \left(\frac{K_p}{2}\right)\theta_{final}.$$

Fig. XX-12. (a) Minimum time plots. (b) Phase trajectory.

Tabulated below is a comparison of the actual eye movement behavior and the minimum time behavior.

Actual Behavior		Minimum Time Behavior	
Movement (degrees)	Time (msec)	Movement (degrees)	Time (msec)
0 to 6.25	35	0-6.6	9.4
0 to 12.5	52	0-13.6	18.8
0 to 18.75	65	0-21.8	32.6
0 to 25	94	0-28.2	43.8
6.25 to 12.5	35	6.3-12.2	9.4
6.25 to 18.75	56	6.3-18.6	18.8
6.25 to 25	73	6.3-26.1	32.6
		12.5-18	9.8
12.5 to 25	60	12.5-24	19.4
18.75 to 25	44	18.75-23.9	10.1

It is evident that the actual behavior of the system requires from 2 to 4 times as much time to execute a movement as would be the case if the system were minimizing time. It would be of interest to compare the actual system behavior with that which would result if other cost functions were minimized. Another consideration is that there may be transient processes between the nerve stimulation and the variables T_{Ag} and T_{Ant} of Fig. XX-5.

G. Cook, L. Stark

References

1. D. A. Robinson, The mechanics of human saccadic eye movement, J. Physiol. 174, 245-264 (1964).

2. Ibid., see Fig. 8, p. 254.

3. D. R. Wilkie, Facts and theories about muscle, Progress in Biophysics and Biophysical Chemistry, Vol. 4, pp. 294; 314, 1954.

4. Ruch and Fulton, Medical Physiology and Biophysics (W. B. Saunders Company, Philadelphia and London), p. 108.

5. A. V. Hill, J. Physiol. 93, 4 (1938).

6. F. H. Adler, Physiology of the Eye (The C. V. Mosby Company, St. Louis, 1959), p. 363.

7. B. Katz, The relation between force and speed in muscular contraction, J. Physiol. 96, 56 (1939).

8. G. M. Breinin, Analytic studies of the electromyogram of human extraocular muscle, Am. J. Ophthalmol. 46, Part II, 131 (September 1958).

9. Inman et al., Relation of human electromyogram to muscular tension, E.E.G. Clin. Neurophysiol. 4, 187-194 (1952).

10. H. Davson, The Eye, Vol. 3 (Academic Press, New York and London, 1962), p. 159.

C. HOSPITAL INFORMATION SYSTEMS: AN AUTOMATED TUMOR REGISTRY

At the Massachusetts Memorial Hospitals of the Boston University Medical Center we have undertaken the design and implementation of an automated Tumor Registry system. It is planned that this system will be readily accessible for administrative, medical, and research use by hospital personnel. The present Tumor Registry is contained in folders and file cabinets and has proved inadequate for the demands of the hospital community. The new system (Fig. XX-13) will be organized around an IBM 1620 digital computer and will use conventional hospital records and the Termatrex Information System. This report describes the organization and programming of an initial pilot system for the Tumor Registry.

The basic components consist of an IBM Model II 1620 computer with 60,000 memory cores, punched card and typewriter input/output, and a 1311 disk storage drive. The computer is equipped with the Monitor I executive routine. The Termatrex System
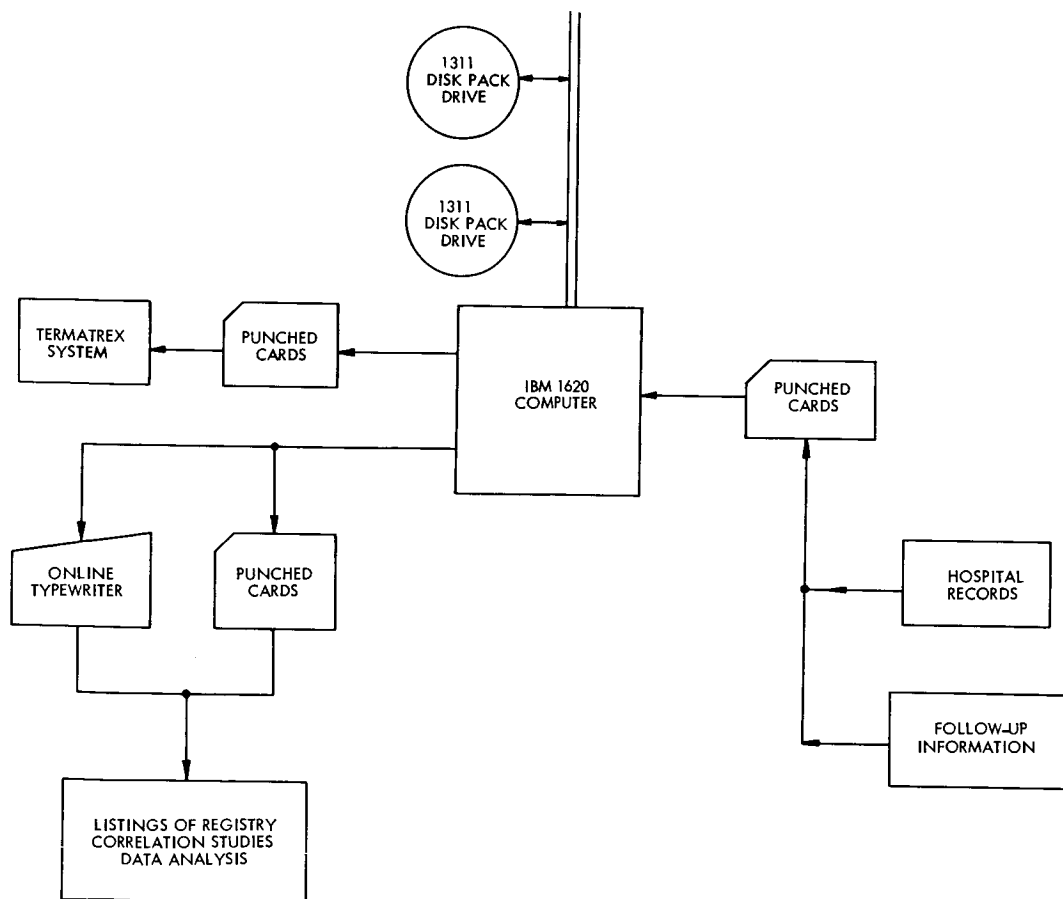

Fig. XX-13. Automated tumor registry system.

is a hand-operated, cross-referenced filing system; by assigning predetermined codes to the various characteristics of a case history, it can accommodate several hundred characteristics of as many as 10,000 patients.

The computer serves as a clearing house and storage area for the Tumor Registry. All information entering the Registry initially passes through the computer, where three types of processing occur. First, the Model 1620 checks the data for errors; for example, a birth date of June, 1967 is an obvious mistake; less obvious errors might be the incorrect spelling of a diagnosis or a nonexistent type of medication. When the data have passed the error-check routine the computer matches each item of the patient's record with the appropriate Termatrex code. The code is then output immediately on punched cards or directly on the 1620's on-line typewriter. The third processing function is storage of the information on the disk pack. This constitutes the actual entry of information into the Registry. Before storage, the data are broken down into two major categories, initial data and subsequent or follow-up data. These categories conform to the pattern of the conventional hospital record and provide a convenient way of organizing the Registry on the disk.

For the pilot system there are approximately 20 terms describing the initial condition of a tumor, and 8 terms for each follow-up report. The initial data are stored at the front of the disk pack and arranged in sequential tables with one table for each term. Follow-up data for each tumor are stored as a pointer list, starting directly after the last entry of the initial data tables. Theoretically, there is no upper limit to the number of tumor cases that can be stored in the completed Registry, as more disk packs can be added to increase its capacity. The pilot system's size, however, is now limited to 1008 cases, as this number can be easily handled on one disk pack, and, at the same time, is large enough to permit a thorough debugging of the system.

The system is augmented by several main programs, corresponding to the major functions of the system, such as data storage, updating, and information retrieval. Figure XX-14 shows the organization of the data-storage program. The main program is a control program; it keeps track of which data has been processed and which subprogram is next to be executed. Each subprogram processes one item of the input record. During execution, control passes from the main program to a subprogram, back to the main program, and then on to another subprogram. The subprograms are completely independent of each other and can be executed in any order and any number of times for each case input. Part of the data of each input record is a code telling the main program the order in which the subprograms are to be executed to process that input record.

This structure permits great flexibility in deciding which characteristics should be recorded in the system. If a new term must be added to each patient's record, a new subprogram is written and added to the rest. Space on the disk pack for storing the new
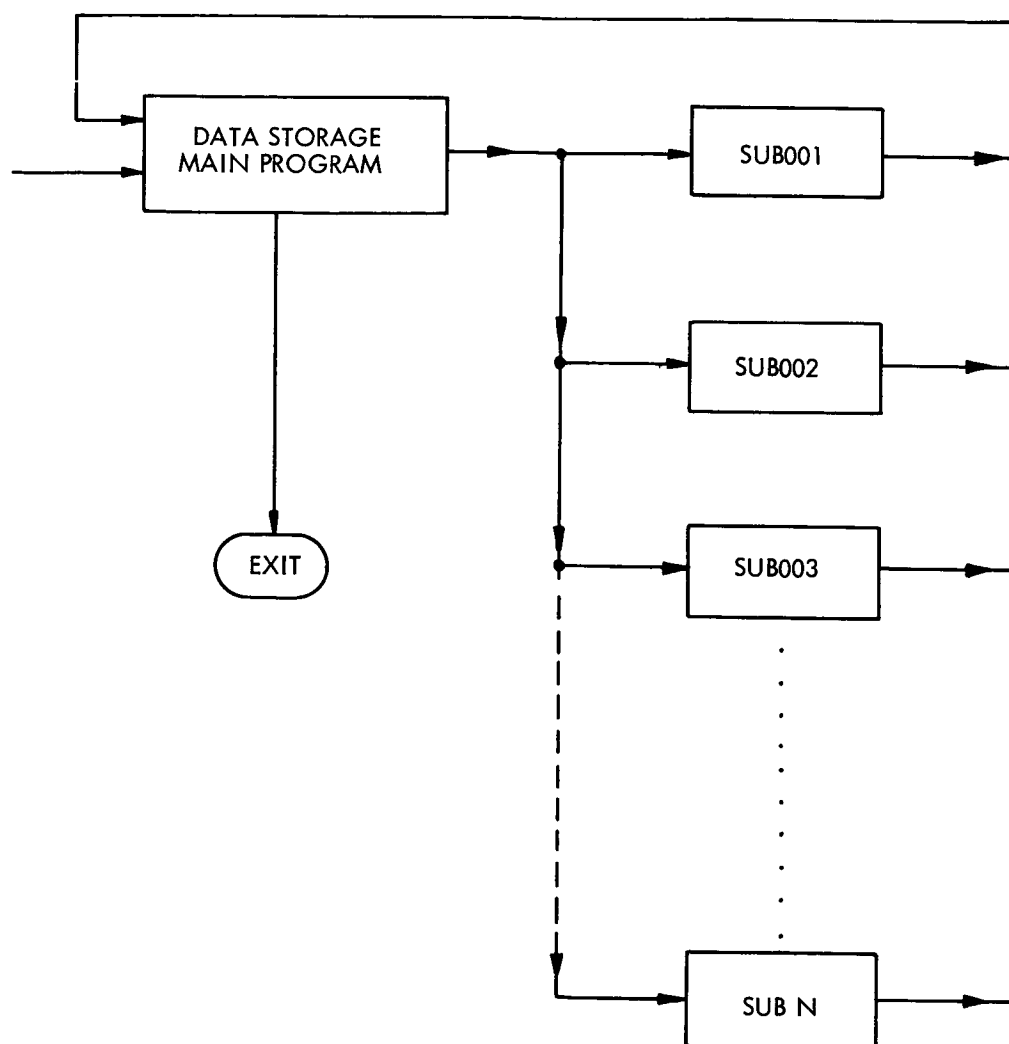
Fig. XX-14. Data-storage programs.

term can be procured either by copying the entire disk onto another disk and leaving an appropriate amount of blank space between the previous last initial data table and the first follow-up data, or by isolating a block of space in the follow-up data area.

The actual method of entering data into the Tumor Registry is still rather primitive. The pertinent data are abstracted from the patient's record, keypunched in a standard abbreviated form, and then introduced into the Model 1620. When the total system is operational, it should be possible to set up a direct communication link from data-collection points to the computer through a time-sharing system teletype; this would permit either an abstracter or a physician to enter the data directly into the Registry. Use of such a system for entering the data would greatly facilitate updating of the Registry.

Information retrieval is, of course, the major goal of a computerized Tumor Registry. For example, the computer can produce, in a very short time, a listing of the entire Registry of the patients who are due for follow-up visits in a given month, or of patients with more than one primary tumor. It readily allows comparisons to be made of the results of a similar treatment of different tumors, of survival times of patients with tumors in various stages of growth or of the frequency with which a clinical diagnosis is confirmed by histological evidence. Because the information is stored essentially by characteristic and not by patient, the search time for a given set of characteristics is minimal.

The work, thus far, has been concerned with laying out the format of the disk pack, comparing various methods of data storage, and coding initial data-storage programs. The disk pack format is being designed for the entire group of initial and follow-up terms, and data-storage subprograms are being coded for all of these terms.

T. Ostrand, J. F. Dickson

D. INTERPRETIVE AND DIAGNOSTIC MATRICES FOR COMPUTER DIAGNOSIS OF ELECTROCARDIOGRAMS

1. Introduction

Figure XX-15 is a block diagram of the operational sequence in a computer system designed for remote on-line, real-time diagnosis of clinical electrocardiograms (EKG). In operation, an executive monitor which is used by a hospital technician for remote control of the sequence of operations, is placed in the G. E. 225 computer in the laboratory of the Neurology Group at the Massachusetts Institute of Technology. The diagnostic process is then begun by the hospital technician with the telephone line transmission of a patient's identifying data and EKG signals to the computer.

The EKG signals originating in the hospital are often obscured by noise, and preprocessing is necessary to facilitate rhythm interpretation, pattern recognition, and parameter extraction. The rhythm section, analyzing the atrial and ventricular rates and prematurities, yields approximately 20 mutually exclusive, tentative, rhythm interpretations. In the morphological identification section of the program, the current filters analyze the X-lead of Frank's orthogonal lead system. The P, QRS, and ST-T segments of the EKG signal are introduced to adaptive matched-filter pattern-recognition programs that provide tentative pattern-recognition interpretations. Following point recognition, the parameter-extracting portion of the program makes pertinent determinations that relate to amplitude (as amplitude of the QRS complex) and interval (as the Q-T interval).
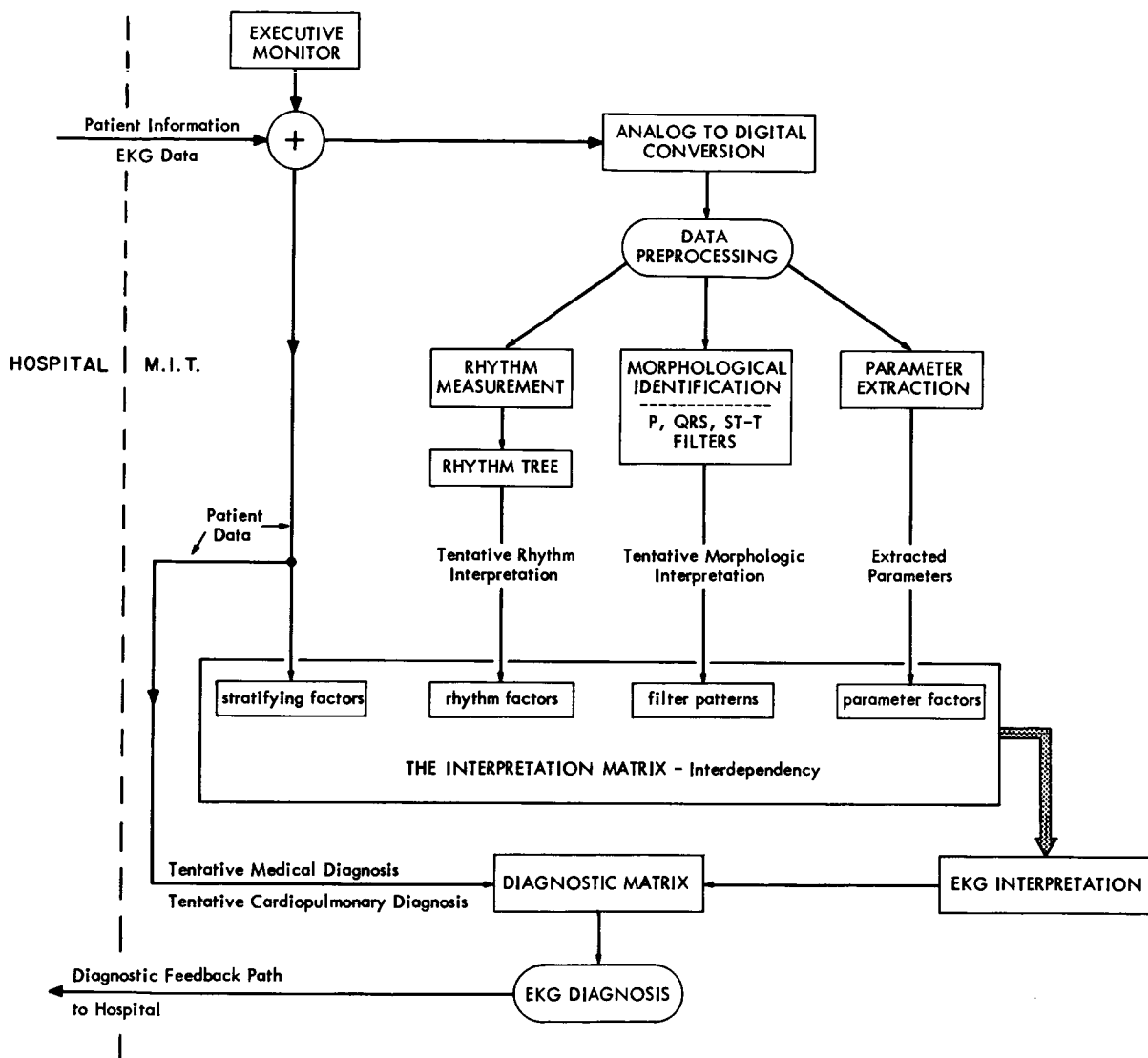
Fig. XX-15. Sequence of the diagnostic system.

2. Interpretive and Diagnostic Matrices

In this part of the diagnostic sequence the independently determined rhythm, pattern recognition, amplitude, and duration parameters are allowed to interrelate for the first time in an additive pattern-recognition matrix. Stratifying factors, such as age, sex, ponderal index, drugs, etc. are also admitted to the diagnostic sequence at this point in the program. Fixed weightings are assigned for the various diagnostic criteria either arbitrarily or adaptively by a learning matrix technique. This matrix yields 25 EKG interpretations (Fig. XX-16), in addition to the rhythm interpretations mentioned above.

1. Within normal limits
2. Left ventricular hypertrophy
3. Right ventricular hypertrophy
4. Complete left bundle branch block
5. Incomplete left bundle branch block
6. Complete right bundle branch block
7. Incomplete right bundle branch block
8. Intraventricular block
9. Acute myocardial infarction
10. Anteroseptal myocardial infarction
11. Inferior myocardial infarction
12. Lateral myocardial infarction
13. Digitalis effect
14. Digitalis intoxication
15. Hyperkalemia
16. Hypokalemia
17. Hypocalcemia
18. Nonspecific ST-T abnormalities
19. Marked ST depression
20. Pericarditis
21. Left atrial hypertrophy
22. Right atrial hypertrophy
23. Acute cor pulmonale
24. Chronic cor pulmonale
25. Wolff-Parkinson-White syndrome

Fig. XX-16.   Twenty-five EKG interpre-
tations contained in the in-
terpretation matrix.

These findings are returned by telephone line for display by teletype at the hospital laboratory.

In the course of reading electrocardiograms one essentially looks for or encounters many information points and then sums their contribution over a decision area for a final diagnosis. We have developed this weighted additive matrix to operate somewhat similarly. It is an adaptive, linear model that accepts nonbinary, probabilistic inputs, does not require that its outputs (diagnoses) be independent, and can display straightforward numerical reasons for its decisions. This matrix now has 174 rows and 25 columns. The rows are divided into 43 categories relating to four types of data: (i) stratified clinical information, such as age, height, weight, sex, and electrocardiographically important drugs being taken, (ii) tentative rhythm analyses, (iii) tentative adaptive matched-filter interpretations, and (iv) the result-ant findings of the parameter extraction program. The columns represent the 25 commonly encountered EKG interpretations that have been selected from a frequency table of diagnoses made in the EKG laboratory of the Massachusetts Memorial Hospitals in the years 1961-1962.

The values used to fill the matrix may range from −500 to +500, a negative value indicating that a particular entry in the matrix weights negatively in the determination of a particular interpretation, as with a short QRS duration for bundle branch block. Zero entries indicate that the information does not contribute to a decision. For the filters the assigned weighted number for pattern recognition is multiplied by the deter-mined correlation coefficient, though it is not clear at this time that the correlation coefficient should necessarily modify the weighted number linearly. Originally, the matrix values were assigned arbitrarily on the basis of clinical experience; however, a self-learning program for the matrix now makes use of our EKG library to refine these values. In the calculation process a patient vector is created, each element of which corresponds to a specific row of the matrix. In each case, except for the filters,

a coefficient of 0 or 1 is assigned as is appropriate. For the filters the correlation coefficient is used. The patient vector is multiplied by the matrix, after which each column is summed to yield 25 additive interpretation totals. Since even the largest sum may represent an unlikely interpretation, the final step is to establish a "certainty" normalization for these interpretation totals. This is done as follows. A maximum possible value for each category is +500. An interpretation of maximum certainty would give a column total equal to the sum of the exhibited patient vector coefficients multiplied by 500 for an ideal sum. Accordingly, the sum of each column is divided by the ideal sum, and yields a certainty factor between minus one and plus one for each interpretation. The interpretation and diagnostic matrices are similarly constructed, so that the final output of the system is an EKG diagnosis with a certainty factor. A means for the automatic adjustment of weightings by a "self-learning" routine is available.

Category	Location	Exhibited Range or Datum	Interpretation Matrix
Age	1	0-20 years	
	2	21-40	
	3	41-60	
	4	over 61	
Sex	5	male	
	6	female	
Height	7	under 60 inches	
	8	60-64	
	9	65-69	
	10	70-75	
	11	over 75	
Weight	12	under 100 lbs.	
	13	101-150	
	14	150-200	
	15	over 200	
Mean Blood Pressure	16	60-80 mm. Hg	
	17	81-100	
	18	101-120	
	19	121-140	
	20	140-160	
QRS filter (pattern) number	21	#1	
	22	#2	
	23	#3	
	24	#4	
	25	#5	

Matrix Output
Right ventricular hypertrophy
Left ventricular hypertrophy
Anteroseptal myocardial infarction
Inferior myocardial infarction
Normal tracing

Fig. XX-17. Abbreviated interpretation matrix.

It reinforces or negatively reinforces appropriate matrix elements, since data from a patient with a known EKG interpretation are passed through, the outcome depending on whether the matrix arrives at a correct or an incorrect interpretation.

The approach just outlined is demonstrated by the following example (Fig. XX-17). The input data for a given patient are entered in the appropriate locations (rows) of a column vector whose potential entries correspond exactly to the named entries (rows) of the interpretation matrix. Most of this information is in the form of graded categories. The rows are labeled with respect to the patient's age, height, mean blood pressure, weight, sex, and type of electrocardiographic pattern.

The matrix on the right side of the figure can yield only five interpretations in this instance (the result has been abbreviated here for pedagogical purposes). When the data on a given patient are entered in the column vector on the left, the corresponding rows of the matrix become operational; that is, the weighting coefficients in the rows to which the patient has contributed data are selected for the diagnostic summing process. For a male patient, 25 years old, whose height is 68 inches, weight 162 lbs, mean blood pressure 98 mm mercury, and whose electrocardiographic QRS complex most closely resembles the pattern of filter No. 2, the rows 2, 5, 9, 14, 17, and 22 of the matrix would become operational. The weighting coefficients located in each cell of these rows would then be summed for each of the 5 diagnostic columns of the matrix. The interpretation assigned to the column with the largest sum would then be output as the most likely interpretation, together with a "certainty coefficient" that would indicate the likelihood of its correctness (in this case the sum of the observed coefficients divided by 12, 500).

In the case of the "learning matrix," each cell in every row and column of the matrix would initially be zero. Consequently, when the very first patient was put through, the sum in each column would be zero and all five diagnoses would be equally possible. For this case, an arbitrary tie-breaking decision has been assigned: the machine is instructed to take the leftmost diagnostic column in case of ties and consider that to be the diagnosis. Our hypothetical first patient would thus be given the diagnosis of "right ventricular hypertrophy" (column 1).

Suppose, however, that his correct diagnosis as established clinically was "normal tracing" (column 5). During the matrix learning process, this correct answer would also have been entered in the computer. A comparison between the correct answer and the answer given by the matrix would be made. In case of a discrepancy, such as that just outlined, all of the operational entries in the right ventricular hypertrophy column would be negatively reinforced by subtracting an arbitrary amount from those weighting coefficients. Similarly, all operational entries in the 5th column would be reinforced by addition of the same amount as was previously subtracted in the "punishing" routine. The "punishment" and "reward" would be equalized in this way, to prevent the numerical value of the matrix as a whole from drifting off-scale. The average value of the

matrix at all times would thus be equal to zero. If this sequence were frequently repeated by processing many patients with known diagnoses, the matrix of weighting coefficients would ideally converge toward the situation wherein the individual entries would have a meaning analogous to conventional statistical discriminant function coefficients.

The advantages of the present "adaptive" scheme are its greater convenience (automaticness) and its open-ended quality. By the last we mean that the stored matrix can be improved at any time without having to destroy the current matrix and start afresh each time a new group of patients with known diagnoses is to be processed. On the other hand, with the learning matrix scheme outlined above, the matrix in its current state can be used to interpret unknown tracings at any time by operating in a fixed, nonlearning form.

For research purposes, the interpretation matrix must contain various indexing and bookkeeping rows and columns that keep track of the types of previously diagnosed electrocardiograms entering into the adaptive matrix, the number of the various clinical input and filter parameters actually utilized, and a tally with respect to the number of correct matrix interpretations and normalizing factors giving estimates of the validity of the weighting coefficients in the matrix at any given time. It will thus be possible to assess quantitatively, perhaps for the first time in extenso, the importance of various clinical and morphologic characteristics for a given interpretation.

<div align="right">J. F. Dickson, D. Martin, G. H. Whipple</div>

XXI. NETWORK SYNTHESIS

Prof. H. B. Lee   J. Andersen   V. K. Prabhu
Prof. W. C. Schwab   R. S. Smith

A. AN ADDITIONAL REALIZATION CYCLE FOR LC IMPEDANCES

Several authors[1,2] recently have introduced new realization cycles for lossless driving-point impedances. The cycles in question develop the impedance function into unsymmetrical lattices terminated in simpler impedances. In the present report we introduce another realization cycle for lossless impedances, one that develops the impedance into an unsymmetrical bridged tee terminated in a simpler impedance.

The new cycle is shown in Fig. XXI-1. The cycle is appropriate to lossless impedances of the form

$$Z(s) = A \frac{s\left(s^2+a_1^2\right) \cdots \left(s^2+a_{n-1}^2\right)}{\left(s^2+\beta_n^2\right) \cdots \left(s^2+\beta_n^2\right)} \tag{1}$$

and has associated a remainder function of the form

$$Z_r(s) = H \frac{s\left(s^2+\gamma_1^2\right) \cdots \left(s^2+\gamma_{n-3}^2\right)}{\left(s^2+\delta_1^2\right) \cdots \left(s^2+\delta_{n-2}^2\right)}. \tag{2}$$

The cycle is canonic, since it achieves a four-coefficient simplification of the driving-point function at a cost of four circuit elements.

The procedure for executing the cycle is as follows.

1. Make two total pole removals from $Z(s)$ at infinity, one on the admittance basis, and one on the impedance basis, as shown in Fig. XXI-2.

2. Extract a Brune section from the remainder function $Z_1(s)$ as shown in Fig. XXI-2; select the null frequency ω_o of the Brune section to be any positive real root of the equation

$$0 = 2\frac{Z_1(j\omega_o)}{j\omega_o} - 2C_o\omega_o^2\left[\frac{Z_1(j\omega_o)}{j\omega_o}\right]^2$$

$$-L_oC_o\omega_o^2\left[Z_1'(j\omega_o) + \frac{Z_1(j\omega_o)}{j\omega_o}\right]. \tag{3}$$

3. Replace the dashed two-port of Fig. XXI-2 by an equivalent two-port of the type shown in Fig. XXI-3. (This can be done conveniently by computing z_{22} for the two-port and then developing $z_{22}(s)$ into a ladder having series capacitors and shunt inductors.)

To establish the validity of the procedure it is only necessary to show that the boxed

two-port of Fig. XXI-2 is equivalent to a two-port of the type shown in Fig. XXI-3, under the condition (3). This can be done as follows.

Direct analysis of the boxed two-port shows that its impedances take the form

$$z_{22} = \frac{ds^4 + cs^2 + b}{s^3 + as} \qquad (4a)$$

$$z_{12} = \frac{es^2 + b}{s^3 + as} \qquad (4b)$$

$$z_{11} = \frac{F(a,b,c,d,e)s^2 + b}{s^3 + as}, \qquad (4)$$

where the parameters a, b, c, d, e are functions of C_o, L_o, L, C, and ρ, and $F(a,b,c,d,e)$ is a function of a, b, c, d, e which makes the $z_{ij}(s)$ compact at $s = \pm j\sqrt{a}$.

The realizability conditions for the two-port of Fig. XXI-3 are readily found[3] to be

  I. The $z_{ij}(s)$ must satisfy the general conditions for lossless realizability.
  II. The $z_{ij}(s)$ must take the form

$$z_{22} = \frac{d's^4 + c's^2 + b'}{s^3 + a's} \qquad (5a)$$

$$z_{12} = \frac{c's^2 + b'}{s^3 + a's} \qquad (5b)$$

$$z_{11} = \frac{F(a',b',c',d',c')s^2 + b'}{s^3 + a's}. \qquad (5c)$$

Clearly, the impedances (4a, b, c) satisfy the foregoing realizability conditions, provided that

$$e = c. \qquad (6)$$

Direct calculation shows that (6) amounts to the requirement

$$0 = \frac{L}{C}(\rho-1)^2 + \frac{\rho(\rho-1)L}{C_o} + \frac{L_o}{C}. \qquad (7)$$
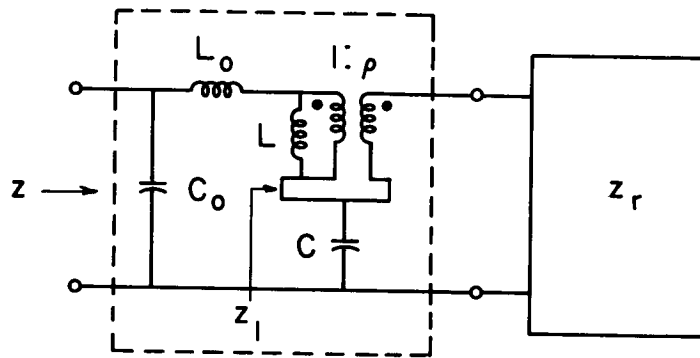
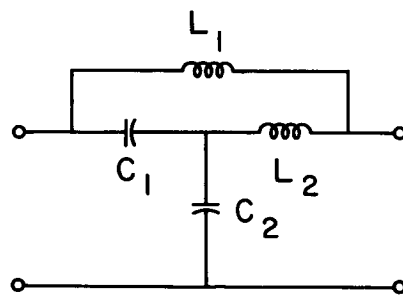Substitution of the well-known[4] values
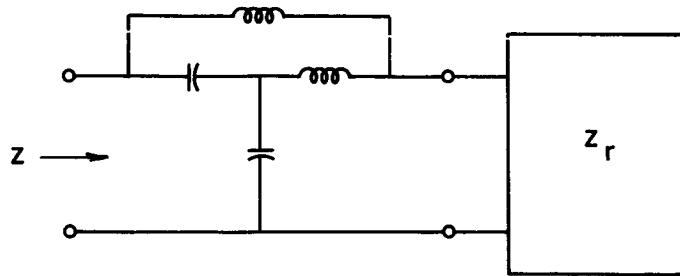
Fig. XXI-1.

Fig. XXI-2.

Fig. XXI-3.

$$L = \frac{1}{2}\left[Z_1'(j\omega_0) + \frac{Z_1(j\omega_0)}{j\omega_0}\right]; \quad C = \frac{2}{\omega_0^2\left[Z_1'(j\omega_0) - \frac{Z_1(j\omega_0)}{j\omega_1}\right]}$$

$$\rho = \frac{Z_1'(j\omega_0) - \frac{Z_1(j\omega_0)}{j\omega_0}}{Z_1'(j\omega_0) + \frac{Z_1(j\omega_0)}{j\omega_0}}$$

and extensive simplification lead to (3). Hence the boxed two-port is equivalent to a two-port of the type shown in Fig. XXI-3, provided that (3) is satisfied.

The following considerations show that (3) admits of at least one positive real solution.

(i) The right-hand member of (3) is a continuous function of ω_0 on the interval $0 < \omega_0 < \omega_1$ (ω_1 denotes the lowest frequency at which $Z_1(j\omega_0)$ has a pole).

(ii) The right-hand member is positive at $\omega_0 = 0 + \left[\approx 2Z_1'(0)\right]$, and is negative at $\omega_0 = \omega_1 - \left[\approx -\left(2C_0 k^2 + L_0 C_0 \omega_0^2 k\right)\frac{1}{(\omega_0 - \omega_1)^2}\right.$, where $k = \text{Res } \left.[Z(j\omega_0)]\right|_{\omega_0 = \omega_1}$.

The basic cycle has several variants. These are as follows.

1. The cycle that results when the capacitors and inductors of Fig. XXI-1 are interchanged.

2. The dual of the cycle shown in Fig. XXI-1.

3. The dual of the first variant.

The first variant is appropriate to impedances of form (1), and has associated a remainder impedance of form (2). The second and third variants are appropriate to impedances of the form

$$Z(s) = A\frac{\left(s^2 + \alpha_1^2\right) \ldots \left(s^2 + \alpha_n^2\right)}{s\left(s^2 + \beta_1^2\right) \ldots \left(s^2 + \beta_{n-1}^2\right)}, \tag{8}$$

and have associated remainder impedances of the form

$$Z_r(s) = H\frac{\left(s^2 + \gamma_1^2\right) \ldots \left(s^2 + \gamma_{n-2}^2\right)}{s\left(s^2 + \delta_1^2\right) \ldots \left(s^2 + \delta_{n-3}^2\right)} \tag{9}$$

H. B. Lee

References

1. H. B. Lee, A new canonic realization procedure, IEEE Trans. on Circuit Theory, Vol. CT-10, pp. 81-85, March 1963.

2. R. Yarlagadda and Y. Tokad, On the use of nonsymmetric lattice sections in network synthesis, IEEE Trans. on Circuit Theory, Vol. CT-11, pp. 474-478, December 1964.

3. These conditions can be established by essentially the same line of reasoning as was used by H. B. Lee, loc. cit., to show that (6a, b, c) constitute realizability conditions for the two-port considered there.

4. D. Hazony, Elements of Network Synthesis (Reinhold Publishing Corporation, New York, 1963), Chap. 8, p. 140.

JOINT SERVICES DISTRIBUTION LIST

Department of Defense

Dr Edward M. Reilley
Asst Director (Research)
Ofc of Defense Res & Eng
Department of Defense
Washington, D.C. 20301

Dr James A. Ward
Office of Deputy Director (Research
and Information Rm 3D1037
Department of Defense
The Pentagon
Washington, D.C. 20301

Director
Advanced Research Projects Agency
Department of Defense
Washington, D.C. 20301

Mr Charles Yost, Director
For Materials Sciences
Advanced Research Projects Agency
Department of Defense
Washington, D.C. 20301

Defense Documentation Center
Attn: TISIA
Cameron Station, Bldg 5
Alexandria, Virginia 22314

Director
National Security Agency
Attn: C3/TDL
Fort George G. Meade, Maryland 20755

Department of the Army

Chief of Research and Development
Headquarters, Department of the Army
Attn: Physical Sciences Division P&E
Washington, D.C. 20310

Research Plans Office
U.S. Army Research Office
3045 Columbia Pike
Arlington, Virginia 22204

Commanding Officer
Foreign Service & Technology Center
Arlington Hall
Arlington, Virginia

Commanding General
U.S. Army Materiel Command
Attn: AMCRD-RS-PE-E
Washington, D.C. 20315

Commanding General
U.S. Army Strategic Communications
Command
Washington, D.C. 20315

Commanding General
U.S. Army Materials Research Agency
Watertown Arsenal
Watertown, Massachusetts 02172

Commanding Officer
U.S. Army Ballistics Research Laboratory
Attn: V.W. Richards
Aberdeen Proving Ground
Aberdeen, Maryland 21005

Commandant
U.S. Army Air Defense School
Attn: Missile Sciences Division, C&S Dept.
P.O. Box 9390
Fort Bliss, Texas 79916

Commanding General
U.S. Army Missile Command
Attn: Technical Library
Redstone Arsenal, Alabama 35809

Commanding General
Frankford Arsenal
Attn: SMUFA-1310 (Dr Sidney Ross)
Philadelphia, Pennsylvania 19137

U.S. Army Munitions Command
Attn: Technical Information Branch
Picatinney Arsenal
Dover, New Jersey 07801

Commanding Officer
Harry Diamond Laboratories
Attn: Mr Berthold Altman
Connecticut Avenue and Van Ness Street N.W.
Washington, D.C. 20438

Commanding Officer
Harry Diamond Laboratories
Attn: Dr R.T. Young
Electron Tubes Division
Connecticut Avenue and Van Ness Street N.W.
Washington, D.C. 20438

Commanding Officer
U.S. Army Security Agency
Arlington Hall
Arlington, Virginia 22212

Commanding Officer
U.S. Limited War Laboratory
Attn: Technical Director
Aberdeen Proving Ground
Aberdeen, Maryland 21005

Commanding Officer
Human Engineering Laboratories
Aberdeen Proving Ground
Maryland 21005

Director
U.S. Army Engineer Geodesy,
   Intelligence and Mapping
Research and Development Agency
Fort Belvoir, Virginia 22060

Commandant
U.S. Army Command and General Staff
   College
Attn: Secretary
Fort Leavenworth, Kansas 66207

Dr. H. Robl, Deputy Director
U.S. Army Research Office (Durham)
P.O. Box CM, Duke Station
Durham, North Carolina 27706

Commanding Officer
U.S. Army Research Office (Durham)
Attn: CRD-AA-IP (Richard O. Ulsh)
P.O. Box CM, Duke Station
Durham, North Carolina 27706

Commanding General
U.S. Army Electronics Command
Attn: AMSEL-SC
Fort Monmouth, New Jersey 07703

Director
U.S. Army Electronics Laboratories
Attn: Dr S. Benedict Levin, Director
   Institute for Exploratory Research
Fort Monmouth, New Jersey 07703

Director
U.S. Army Electronics Laboratories
Attn: Mr Robert O. Parker, Executive
   Secretary JSTAC (AMSEL-RD-X)
Fort Monmouth, New Jersey 07703

Superintendent
U.S. Army Military Academy
West Point, New York 10996

The Walter Reed Institute of Research
Walter Reed Army Medical Center
Washington, D.C. 20012

Director
U.S. Army Electronics Laboratories
Fort Monmouth, New Jersey 07703
Attn: AMSEL-RD-DR   NE   SS
            X   NO   PE
            XE   NP   PR
            XC   SA   PF
            XS   SE   GF
            NR   SR   ADT
                  FU#1

Commanding Officer
U.S. Army Electronics R&D Activity
Fort Huachuca, Arizona 85163

Commanding Officer
U.S. Army Engineers R&D Laboratory
Attn: STINFO Branch
Fort Belvoir, Virginia 22060

Commanding Officer
U.S. Army Electronics R&D Activity
White Sands Missile Range
New Mexico 88002

Director
Human Resources Research Office
The George Washington University
300 N. Washington Street
Alexandria, Virginia 22314

Commanding Officer
U.S. Army Personnel Research Office
Washington, D.C.

Commanding Officer
U.S. Army Medical Research Laboratory
Fort Knox, Kentucky

Department of the Air Force

Director
Air University Library
Maxwell A.F. Base, Alabama

Commander
Air Force Office of Scientific Research
Washington 25, D.C.
Attn:-SREE

Department of The Air Force
Headquarters-United States Air Force
Washington 25, D.C.
Attn: AFTAC/TD-1

Dr. Harvey E. Savely, SRL
Air Force Office of Sci. Res.
Office of Aerospace Research, USAF
Washington 25, D.C.

Mr. C.N. Hasert
Scientific Advisory Board
Hq, USAF
Washington 25, D.C.

APGC (PGBAP-1)
Elgin Air Force Base
Florida 32542

AFETR
(AFETR Tech. Library MU-135)
Patrick Air Force Base
Cocoa, Florida

Air Force Cambridge Res. Lab.
L.G. Hanscom Field
Bedford, Massachusetts 01731
Attn: CRDM, Mr. Herskovitz

Commander, AFCRL
Attn: C.P. Smith (CRBS)
L.G. Hanscom Field
Bedford, Massachusetts

Dr. L. C. Block
AFCRL (CROV)
L. G. Hanscom Field
Bedford, Massachusetts

AFCRL
Office of Aerospace Res., USAF
Bedford, Mass.
Attn: CRDA

Mr. Rocco H. Urbano, Chief
AFCRL, Appl. Math. Branch
Data Sciences Laboratory
Laurence G. Hanscom Field
Bedford, Massachusetts

AFCRL (CRFE-Dr. Nicholas Yannoni)
L.G. Hanscom Field
Bedford, Massachusetts

S. H. Sternick
Aerospace Comm. - Attn: ESNC
Waltham Federal Center
424 Trapelo Road
Waltham, Massachusetts 02154

Rome Air Dev. Center (RAWL, H.Webb)
Griffiss Air Force Base
New York 13442

Systems Engineering Group
Deputy for Systems Eng'g., SEPRR
Directorate of Tech. Pubs. and Specs.
Wright-Patterson AFB, OHIO 45433

Aeronautical Systems Division
Attn: ASRPE, Mr. Robt. Cooper
Wright-Patterson AFC, Ohio 45433

Aeronautical Systems Division
Attn: ASRPP-20 (Mr. Don R. Warnock)
Wright-Patterson AFB, Ohio 45433

AFAL
AVR (L)
Wright-Patterson AFB
Ohio 45433

Dr. H. H. Kurzweg
Director Research - OART
NASA
Washington, D.C. 20546

Systems Engineering Group (RTD)
Attn: SEPIR
Wright-Patterson AFB, Ohio 45433

AFAL (AVTE)
Wright-Patterson AFB
Ohio 45433

Mr. Roland Chase
National Aeronautics & Space Administration
1512 H Street, N.W.
Washington 25, D.C.

Professor Arwin Dougal
University of Texas
EE Department
Austin, Texas

Honorable Alexander H. Flax
Asst Secretary of the Air Force (R&D)
Office of the Secretary of the Air Force
Washington 25, D.C.

Professor Nicholas George
California Institute of Technology
EE Department
Pasadena, California

Dr. Lowell M. Hollingsworth
AFCRL
L.G. Hanscom Field
Bedford, Massachusetts

Dr. Zohrab Kaprielian
University of Southern California
University Park
Los Angeles 7, California

Dr. John M. Ide
National Science Foundation
Washington 25, D.C.

Lt Col Edwin M. Myers
Headquarters USAF (AFRDR)
Washington 25, D.C.

Professor Wm. H. Radford
Director, Lincoln Laboratories
Lexington, Massachusetts

Brig Gen B.G. Holzman, USAF (Ret.)
Electronics Research Center, NASA
30 Memorial Drive
Cambridge, Mass.

Dr. R. L. Sproull
Director, Advanced Research Projects
    Agency
Washington 25, D.C.

Brigadier General J. T. Stewart
Director of Science & Technology
Deputy Chief of Staff (R&D)
USAF
Washington 25, D.C.

Mr. James Tippett
National Security Agency
Fort Meade, Maryland

Dr. H. Harrison
NASA (Code RRE)
Fourth and Independence Streets
Washington, D.C. 20546

AEC
Civ of Tech Info Ext
P.O. Box 62
Oak Ridge, Tenn.

AFRST (SC/EN)
Lt Col L. Stone
Rm 4C 341
The Pentagon
Washington, D.C. 20301

U.S. Atomic Energy Commission
Library
Gaithersburg, Md. 20760

ARL (ARD/Col R.E. Fontana)
Wright-Patterson AFB,
Ohio 45433

Office of Research Analyses
Attn: Col K.W. Gallup
Holloman AFB, NMex 88330

AFCRL (CRXL)
L.G. Hanscom Fld
Bedford, Mass 01731

Frank J. Seiler Rsch Lab
Library
USAF Academy, Colo 80840

ARL (AROL)
Wright-Patterson AFB,
Ohio 45433

Office of Research Analyses
Library
Holloman AFB, NMex 88330

LOOAR (Library)
AF Unit Post Office
Los Angeles, Calif 90045

Churchill Research Range
Library
Fort Churchill
Manitoba, Canada

Los Alamos Scientific Lab
Attn: Technical Library
Los Alamos, NMex 87544

Battelle Memorial Institute
Technical Library
505 King Avenue
Columbus, Ohio 43201

John Crerar Library
35 West 33rd St.
Chicago, Ill.

Linda Hall Library
5109 Cherry St.
Kansas City, Mo.

National Science Foundation
Library
1951 Constitution Ave., N.W.
Washington, D.C. 20550

Johns Hopkins University
Applied Physics Lab Library
White Oak
Silver Spring, Md. 20910

Stanford Research Institute
Library
820 Mission St.
South Pasadena, Calif. 91030

Southwest Research Institute
Library
8500 Culebra Road
San Antonio, Texas

ARPA, Tech Info Office
The Pentagon
Washington, D.C. 20301

DDR&E (Tech Library)
Rm 3C 128
The Pentagon
Washington, D.C. 20301

Industrial College of the
  Armed Forces
Attn: Library
Washington, D.C.

AFIT (MCLI)
Tech Library
Wright-Patterson AFB
Ohio 45433

AUL 3T-9663
Maxwell AFB, Ala 36112

USAFA (DLIB)
USAF Academy, Colorado 80840

AFSC (Tech Library)
Andrews AFB
Washington, D.C. 20331

ASD (Tech Library)
Wright-Patterson, AFB
Ohio 45433

BSD (Tech Library)
Norton AFB, Calif 92409

ESD (ESTI)
L. G. Hanscom Field, F172
Bedford, Mass 01731

RTD (Tech Library)
Bolling AFC, D.C. 20332

AFFTC (Tech Library)
Edwards AFB, Calif 93523

AFMDC (Tech Library)
Holloman AFB, NMex 88330

AFWL (WLIL, Tech Library)
Kirtland AFB, NMex 87117

APGC (Tech Library)
Eglin AFB, Fla 32542

AEDC (Tech Library)
Arnold AFS, Tenn 37389

RADC (Tech Library)
Griffiss AFB, N.Y. 13442

Director
National Aeronautical Establishment
Ottawa, Ontario, Canada

CIA
OCR/LY/IAS
IH 129 Hq
Washington, D.C. 20505

National Defense Library
Headquarters
Ottawa, Ontario, Canada

Technical Library
White Sands Missile Range
NMex 88002

NASA/AFSS/1 FOB6
Tech Library, Rm 60084
Washington, D.C. 20546

Space Systems Division
Los Angeles Air Force Station
Air Force Unit Post Office
Los Angeles, California 90045
Attn: SSSD

U.S. Regional Science Office/LAOAR
U.S. Embassy
APO-676
New York, N.Y.

Ames Rsch Center (NASA)
Technical Library
Moffett Field, Calif 94035

High Speed Flight Center (NASA)
Technical Library
Edwards AFB, Calif 93523

Goddard Space Flight Center (NASA)
Greenbelt, Md. 20771

Geo. C. Marshall Space Flight
   Center (NASA)
Redstone Arsenal, Ala 35808

Lewis Research Center (NASA)
Technical Library
21000 Brookpark Road
Cleveland, Ohio

Aerospace Corp (Tech Library)
P.O. Box 95085
Los Angeles, Calif 90045

Rand Corporation
1700 Main St.
Santa Monica, Calif 90401

Carnegie Institute of Technology
Science & Engineering Hunt Library
Schenley Park
Pittsburgh, Pa. 15213

California Institute of Technology
Aeronautics Library
1201 East Calif St.
Pasadena 4, Calif

AVCO Research Lab
Library
2385 Revere Beach Parkway
Everett, Mass 02149

Dr. G. E. Knausenberger
c/o Hq. Co. Munich Post
APO 09407
New York, N. Y.

Commander
Space Systems Division (AFSC)
Office of the Scientific Director
Inglewood, California

Commander
Aerospace Systems Division
AFSC
Office of the Scientific Director
Wright-Patterson AFB, Ohio ·

Commander
Aerospace Research Laboratories (OAR)
Office of the Scientific Director
Wright-Patterson AFB, Ohio

Commander
Air Force Cambridge Research Laboratories
Office of the Scientific Director
L. G. Hanscom Field
Bedford, Massachusetts

Commander
Air Force Systems Command
Office of the Chief Scientist
Andrews AFB, Maryland

Commander
Research & Technology Division
AFSC
Office of the Scientific Director
Bolling AFB 25, D.C.

Commander
Rome Air Development Center
AFSC
Office of the Scientific Director
Griffiss AFB, Rome, New York

Department of the Navy

Dr. Arnold Shostak, Code 427
Head, Electronics Branch
Physical Sciences Division
Department of the Navy
Office of Naval Research
Washington, D.C. 20360

Chief of Naval Research, Code 427
Department of the Navy
Washington, D.C. 20360

Chief, Bureau of Weapons
Department of the Navy
Washington, D.C. 20360

Chief, Bureau of Ships
Department of the Navy
Washington, D.C. 20360
Attn: Code 680

Commander
U.S. Naval Air Development Center
Johnsville, Pennsylvania
Attn: NADC Library

Library
U.S. Navy Electronics Laboratory
San Diego, California 92152

Commanding Officer
U.S. Navy Underwater Sound Laboratory
Ft Trumbull
New London, Connecticut

Director
Naval Research Laboratory
Washington, D.C. 20390

Commanding Officer
Office of Naval Research Branch Office
Navy 100, Fleet P.O. Box 39
New York, New York

Chief of Naval Operations
Pentagon OP 07T
Washington, D.C.

Commanding Officer
Officer of Naval Research Branch Office
495 Summer Street
Boston, Massachusetts 02110

Commander
Naval Ordnance Laboratory
White Oak, Maryland
Attn: Technical Library

U.S. Navy Post Graduate School
Monterey, California
Attn: Electrical Engineering Department