UNCLASSIFIED

AD 414870

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA

UNCLASSIFIED

414870

QUARTERLY PROGRESS REPORT

No. 70

JULY 15, 1963

TISIA D

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

CAMBRIDGE, MASSACHUSETTS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS

QUARTERLY PROGRESS REPORT No. 70

July 15, 1963

Submitted by: H. J. Zimmermann

G. G. Harvey

TABLE OF CONTENTS

CONTENTS

CONTENTS

CONTENTS

CONTENTS

PERSONNEL

Administration

Professor H. J. Zimmermann, Director
Professor G. G. Harvey, Associate Director
Mr. R. A. Sayers, Personnel and Business Manager

Advisory Committee

Dean G. S. Brown (Ex officio)
Professor W. W. Buechner
Professor P. Elias
Professor G. G. Harvey
Professor A. G. Hill
Professor I. W. Sizer
Professor H. J. Zimmermann
(Chairman)

Faculty

Allis, W. P. (Absent)
Arden, D. N. (Absent)
Barrett, A. H.
Bekefi, G.
Bers, A.
Billman, K. W.
Bitter, F.
Bose, A. G.
Brazier, Mary A. B.
 (Visiting)
Brown, S. C.
Carabateas, E. N.
Chomsky, A. N.
Chorover, S. L.
Chu, L. J.
Cochran, J. F.
Dennis, J. B.
Dupree, T. H.
Eden, M.
Edgerton, H. E.
Elias, P.
Fano, R. M.
Fodor, J. A.
Freedman, S. I.
Gallager, R. G.
Garland, C. W.
Getty, W. D.
Goldstein, M. H., Jr.
Graham, J. W.
Green, D. M.
Guillemin, E. A.
Gyftopoulos, E. P.
Halle, M.
Hammes, G. G.
Harvey, G. G.
Hatsopoulos, G. N. (Absent)
Haus, H. A.

Heinz, J. M.
Heiser, W. H.
Hennie, F. C., III
Hill, A. G. (Absent)
Hoffman, M. A.
Ingard, K. U.
Jackson, W. D.
Jacobs, I. M.
Jakobson, R.
Javan, A.
Katz, J. J.
Kerrebrock, J. L.
King, J. G.
Kinsey, J. L.
Klima, E. S.
Kyhl, R. L.
Lee, Y. W.
Lettvin, J. Y.
Lidsky, L. M.
Manders, A. M.
Mason, S. J.
Matthews, G. H.
Maxwell, E. (Visiting)
Meissner, H. P.
Melcher, J. R.
Melzack, R.
Minsky, M. L.
More, T., Jr.
Nottingham, W. B.
Oates, G. C.
Peake, W. T.
Penfield, P. L., Jr.
Penhune, J. P.
Perry, C. H.
Pfeiffer, R. R.
Postal, P.
Putnam, H.
Rafuse, R. P.

Reiffen, B. (Visiting)
Reynolds, J. M.
Rose, D. J.
Rosenblith, W. A.
Schetzen, M.
Schreiber, W. F.
Schwab, W. C.
Searle, C. L.
Shannon, C. E.
Shapiro, A. H.
Siebert, W. McC.
Smith, J. L., Jr.
Smullin, L. D.
Stevens, K. N.
Stickney, R. E.
Stockham, T. G., Jr.
Strandberg, M. W. P.
Swets, J. A.
 (Absent)
Teager, H. M.
Teuber, H. L.
Thornton, R. D.
Tisza, L. (Absent)
Townes, C. H.
Troxel, D. E.
Van Trees, H. L., Jr.
Wall, P. D.
Warren, B. E.
Waugh, J. S.
Weiss, T. F.
Whitehouse, D. R.
Whitney, W. M.
Woodson, H. H.
Wozencraft, J. M.
Young, L. R.
Zacharias, J. R.
Zimmermann, H. J.

PERSONNEL

Instructors

Alter, R.
Bohacek, P. K.
Briggs, R. J.
Bruce, J. D.
Buka, Magda
Cooper, R. S.
Crystal, T. H. (Absent)
Dean, L. W., III
East, D. A.

Hall, J. L., II
Huang, T. S.
Huibonhoa, R.
Ingraham, J. C.
Katona, P. G.
Kellner, W. G.
Kincaid, T. G.
Kliman, G. B.
Koskinen, M. F.
Krishnayya, J. G.

Landsman, E.
Liu, C. L.
MacDonald, J. S.
Oppenheim, A. V.
Parente, R. B.
Perlmutter, D. M.
Pierson, E. S.
Schneider, A. J.
Shavit, A.

Lecturers

Pitts, W. H.
Rines, R. H.

Rose, G. A. (Visiting)
Stroke, G. W.

Research Associates

Barlow, J. S.
Bishop, P. O.
Bradley, L. C., III
Burke, J. J., Jr.
Cavaggioni, A.

Cerrillo, M. V.
Dupress, J.
Fiocco, G.
Gerstein, G. L.
Gesteland, R. C.
Hermann, H. T.

House, A. S.
Kilmer, W. L.
Stark, L.
Thompson, E.
Yngve, V. H.

Staff Members

Anderson, A. T., III
Arndt, R. W.
Badessa, R. S.
Bates, V. J.
Blinn, J. C., III
Bosche, Carol M.
Brown, R. M.
Butterfield, B. O.
Campbell, Elizabeth J.
Charney, Elinor K.
Clark, W. A., Jr. (Lincoln
Laboratory)
Congleton, Ann
Crist, A. H.
Darlington, J. L.
Duffy, D. F.
Edwards, D. J.
Farley, B. G. (Lincoln
Laboratory)
Ferretti, E. F.
Foley, J. A., Jr.
Garber, M.
Giberman, E.
Goodall, M. C.
Greibach, Rita J.

Harman, G.
Hewitt, J. H.
Howland, B. H. (Lincoln
Laboratory)
Ingersoll, J. G.
Kannel, Muriel
Kerllenevich, N.
Keyes, R. V., Jr.
Kiang, N. Y-S.
Kierstead, J. D.
Kuiper, J. W.
Kusse, B. R.
Lakoff, G. P.
Levin, M. I.
Lightner, T. M.
Lindsley, V. Susan
Maling, G. C., Jr.
McCarthy, J. J.
McCulloch, W. S.
Mandics, P. A.
Marks, K. M.
Millechia, R. J.
Mulligan, W. J.
O'Brien, F. J.
O'Rourke, Ann M.
Park, Beverly R.

Pennell, Martha M.
Percival, W. K.
Pitts, W. H.
Rao, B. D. N.
Rosebury, F.
Ryan, L. W.
Saltalamacchia, A. J.
Sandel, T. T. (Lincoln
Laboratory)
Satterthwait, A. C.
Sayers, R. A.
Shiffman, C. A.
Smith, P. L.
Steinbrecher, D. H.
Stroke, H. H.
Svihula, G. F.
Szoke, A.
Thomas, Helen L.
Vidale, Eda B.
Viertel, J. J.
Whitney, R. E.
Willis, P. A.
Wolfberg, M. S.
Wright, J. S.
Yaqub, M.

PERSONNEL

Visitors

Arbib, M. A.
Baker, F. H.
 (Postdoctoral Fellow
 of the National Insti-
 tutes of Health)
Bendall, L. K.
 (Postdoctoral Fellow –
 Social Science Research
 Council Study Fellow)
Chihara, C. S.
Fohl, T.
Gross, C.
 (Postdoctoral Fellow
 of the National Insti-
 tutes of Health)
Hall, R. D.
 (Postdoctoral Fellow
 of the National Insti-
 tute of Mental Health)

Imai, N.
Isami, Y. (Fulbright
 Research Scholar in
 Linguistics)
Keyser, S. J.
Kornacker, K.
 (Postdoctoral Fellow
 of the National Insti-
 tutes of Health)
Menyuk, Paula
 (Postdoctoral Fellow
 of the National Insti-
 tutes of Health)
Molnar, C. E.
Parks, J. H.
Rojas, J. A.
 (Postdoctoral Fellow
 of the National Insti-
 tutes of Health)

Rosenbloom, D.
Schiller, P. H.
 (Postdoctoral Fellow
 of the National Insti-
 tutes of Health)
Sellin, K. G.
 (UNESCO Fellow)
Siatkowski, Z.
Szekely, G.
 (Postdoctoral Fellow –
 Rockefeller Foundation)
Troelstra, A.
 (NATO Science
 Fellow)
Walker, D. E.
Walter, W. T.
Wu, K. (R. L. E.
 Postdoctoral Fellow)
Yasui, S.

Graduate Students

Algazi, V. R.
Allen, R. J.
Alvarez de Toledo, F.
Anderson, T. M.
 (USN Fellow)
Andrews, J. M., Jr.
Andrews, M. L.
Appel, E. L.
Arunasalam, V.
Axelrod, F. S.
Badrawi, M. T.
Bar-David, I.
Bartsch, R. R.
Bauer, R. F.
Berlekamp, E. R.
 (National Science
 Foundation Fellow)
Bever, T. G.
Black, W. L.
Blum, M.
Bobrow, D. G.
Botha, D. G.
Brabson, B. B.
 (National Science
 Foundation Fellow)
Brassert, W. L.
Breon, R. K.
 (Teaching
 Assistant)

Brown, J. E. (Porter
 Fellow of the American
 Physiological Society)
Brown, T. S.
Buntschuh, C. D.
Burns, S. K.
 (Teaching Assistant)
Bush, A. M.
Cahlander, D. A.
Capranica, R. R.
 (Communications
 Development Training
 Program Fellow of
 Bell Telephone Labo-
 ratories, Inc.)
Chandra, A. N.
Chapman, J. C.
Chiao, R. Y.
Clark, J. F.
Clayton, R. J.
Clemens, J. K.
 (Teaching Assistant)
Coccoli, J. D.
Coggins, J. L.
Cornew, R. W.
Crowley, J. M.
 (U. S. AEC Fellow)
Cunningham, J. E.
Dethlefsen, R.

Diamond, B. L.
Domen, J. K.
Drouet, M. G. A.
Dym, H.
Dyron, Frances M.
Ebert, P. M.
 (National Science
 Foundation Fellow)
Ellis, J. R., Jr.
Falconer, D. D.
Fetz, E. E.
Flannery, D. L.
 (General Motors
 Fellow)
Fleury, P. A.
Forney, G. D., Jr.
Fraim, F. W., IV
Fraser, J. B.
Fukomoto, A.
Gadzuk, J. W.
Garmire, Elsa M.
Garosi, G. A.
Gaut, N. E. (NASA Fellow)
Gentle, K. W.
Gerry, E. T.
Gerstmann, J.
 (National Science
 Foundation Fellow)
Ghosh, S. K.

Graduate Students (continued)

Glenn, W. H., Jr.
Golub, R.
Goodman, L. M.
(Sloan Fellow)
Gothard, N.
Gramenopoulos, N.
Gray, C. E.
(Teaching Assistant)
Gray, P. R.
(Teaching Assistant)
Grayzel, A. I.
Greata, J. M.
(Teaching Assistant)
Greenblatt, S.
Gronemann, U. F.
Guinan, J. J., Jr.
Guttrich, G. L.
Hall, Barbara C.
(National Science
Foundation Fellow)
Halstead, F. C.
Halverson, W. D.
Hart, T. P.
Hartenbaum, B. A.
Hegblom, E. R.
Henke, W. L.
(National Science
Foundation Fellow
Heywood, J. B.
Holsinger, J. L.
Hooper, E. B., Jr.
Houk, J. C., Jr.
(IBM Fellow)
Hsi, C-F. G.
Hsieh, H. Y.
Ivie, E. L. (National Science
Foundation Fellow)
Iwadare, Y.
Jameson, P. W.
Johnston, W. D., Jr.
Kachen, G. I., Jr.
(Teaching Assistant)
Kamen, E. L.
Kellen, P. F.
Ketterer, F. D.
Kiparsky, R. P. V.
Kniazzeh, A. G. F.
Kocher, D. G.
Koerber, K. C.
Kramer, A. J.
(Teaching Assistant)
Kronquist, R. L.
Kukolich, S. G.
Kuroda, S. Y.
(Whitney Fellow)
Landowne, D.

Langbein, D.
Langendoen, D. T.
Larson, E. L.
Lee, K. S.
Lercari, R. F.
Lerman, S. H.
Levine, R. C.
(Teaching Assistant)
Levison, W. H.
Lewis, A. T.
Lewis, T. B.
Liégeois, F. A.
Lim, S. T.
Little, R. G.
Llewellyn-Jones, D. T.
Lontai, L. N.
Lowry, E. S.
Luckham, D. C.
Lutz, M. A.
Macomber, J. D.
Mark, R. G.
Marks, W. C.
(Teaching Assistant)
Mayo, J. W.
McClees, H. C., Jr.
McMorris, J. A., II
Mendell, L.
Mendelson, J.
Mermelstein, P.
Merrill, E. G.
Miller, S. W.
Morse, D. L.
Naves, F.
Nelson, D. E.
Nolan, J. J., Jr.
Nordgren, A. C.
Nowak, R. T.
Oder, R. R.
Olsen, J. H.
Omura, J. K.
Paul, A. P.
Pauwels, H. (Belgium-
American Educational
Foundation Fellow)
Petrie, L. M., Jr.
(Teaching Assistant)
Pfaff, D. W.
Pilla, M. A.
(Teaching Assistant)
Prabhu, V. K.
Praddaude, H. C.
Pruslin, D. H.
Puri, S.
(Teaching Assistant)
Qualls, C. B.
Rahmani, Mahin

Redi, O.
Reid, M. H.
Reisman, S. S.
Renard, R. H.
Reznek, S. R.
Rizzo, J.
Robinson, C. G.
Rogoff, G. L.
Rook, C. W., Jr.
Rosenbaum, P. S.
Rowe, A. W.
Rummler, W. D.
Sachs, M. B.
(Sloan Fellow)
Sandberg, A. A.
Savage, J. E.
Schlossburg, H.
Schneider, H. M.
Schuler, C. J., Jr.
Serafim, P. E.
Shah, J. C.
Shimony, U.
Simpson, J. I.
Sklar, J. R.
Smith, W. W.
Smythe, D. L., Jr.
(Teaching Assistant)
Snow, K. D.
(Teaching Assistant)
Sobel, I.
Solbes, A.
Spangler, P. S.
Staelin, D. H.
Stage, D. E.
Stanten, S. F. (Hughes
Aircraft Fellow)
Steele, D. W.
Stiglitz, I. G.
Sutherland, W. R.
(National Science
Foundation Fellow)
Taub, A. (National
Institutes of Health
Fellow)
Tepley, N.
Theodoridis, G.
Thiele, A. A.
Thomae, I. H.
(National Science
Foundation Fellow)
Thomas, L. C.
(National Science
Foundation Fellow)
Thornburg, C. O., Jr.
Tomlinson, W. J., III
Topaz, J. M.

Graduate Students (continued)

Tretiak, O. J.
Tse, F. Y-F.
Tulenko, J. S.
  (U.S. AEC Fellow)
Ulrich, P. B.
Vanderweil, R. G.
Van Horn, E. C., Jr.
  (Teaching Assistant)
Wade, C. G.
Wagner, C. E.
  (Teaching Assistant)
Waletzko, J. A.
Wasserman, G. S.
Weingrad, J. S.
Welch, J. R.
Whitman, E. C.

Wickelgren, G. L.
Wiederhold, M. L.
Wilde, D. U.
  (Teaching Assistant)
Wilde, G. R.
Wilensky, S.
Williamson, R. C.
Willke, H. L., Jr.
Wilson, G. L.
  (Teaching Assistant)
Wilson, W. J.
  (Teaching Assistant)
Winett, J. M.
Wissmiller, J. C.
  (National Science
  Foundation Fellow)

Witting, H. L.
  (U.S. AEC Fellow)
Witting, J. M.
  (I.T.T. Fellow)
Witzke, K. G.
Wolf, R. P.
Woo, J. C.
Wright, B. L.
Yarnell, C. F.
Yudkin, H. L.
Zeiger, H. P.
Zuber, B. L.
Zuckerman, B. M.
Zwicky, A. M., Jr.
  (National Defense
  Education Act Fellow)

Undergraduates

Aaron, A. M.
Amdur, N. J.
Aponick, A. A.
Bayles, R. U.
Brincko, A. J.
Brody, Juliana E.
Byers, F. H.
Chang, G. D. Y.
Chang, K.
Colodny, Susan F.
Davis, P.
Deininger, C. C.
DerSarkisian, S. F.
Eggers, T. W.

Elliot, J. L.
Gayle, D. D.
Gibson, E. J.
Goldstein, D. E.
Grabowski, R. E.
Lindes, P.
Moxon, E. C.
Murray, M. J.
Naqui, A. A.
Nelson, G. P.
Norris, R. B., Jr.
O'Halloran, W. F., Jr.
Parchesky, J.

Perkinson, J. C.
Politzer, P. A.
Rome, J. A.
Roy, B.
Rush, R. D.
Scott, T. A.
Slater, T. S.
Smith, D. F.
Sparks, C. E.
Stark, M. F.
Thompson, G. D., Jr.
Wan, A. C. M.
Weidner, M. Y.
Yap, B. K.

Senior Thesis Students

Bender, M. H.
Cannon, J. F.
DiGregorio, R. F.
Epstein, M. R.
Evans, S. A.
Fiory, A. T.

Flicker, J. K.
Hadden, W. J., Jr.
Hassan, A. R.
Lennon, W. J.
Manheimer, W. M.
Mudama, E. L.

Ng, L. C.
Okereke, S. A.
Pearlman, M. R.
Porter, R. P.
Weintraub, A. C.
Winsor, N. K.

Assistants and Technicians

Aquinde, P.
Arnold, Jane B.
Babcock, E. B.
Barrett, J. W.
Barrows, F. W.
Beaton, Catherine M.
Bella, C. J.
Berg, A. E.
Butler, R. E., Jr.

Chase, Arbella P.
Connolly, J. T.
Cook, J. F.
Cranmer, R. E.
Crist, F. X.
DiPietro, P. J.
Donahue, P. D.
Engler, R. R.
Fitzgerald, E. W., Jr.

Footnick, Rosalie
French, Marjorie A.
Gay, H. D.
Grande, Esther D.
Greenwood, E. L.
Gregor, C. A.
Hallett, J. G.
Henning, G. B.
Karas, P.

PERSONNEL

Assistants and Technicians (continued)

Kelly, M. A.
Kirkpatrick, Patty I.
Lewis, R. R.
Major, Diane
Massey, L. N.
McKenzie, J. A.
McLean, J. J.
Misail, M. L.
Molin, A. H.
Neal, R. W.
North, D. K.

Overslvizen, T.
Papa, D. C.
Peck, J. S.
Pyle, Cynthia M.
Rowe, T. A.
Samson, P. R.
Schwabe, W. J.
Sears, A. R.
Shane, Carolyn S.
Shuman, Susanne

Smith, Gabriella W.
Sonnenberg, S. B.
Sprague, L. E.
Stroud, Marion B.
Thompson, D. C.
Tortolano, A. J.
T'sou, B. K-Y.
Vance, H. T., Jr.
Volkmann, Ruth
Yaffee, M. A.
Yee, F. Q.

Document Room

Hewitt, J. H.
Crayton, Marjorie

Fournier, Lorraine E.
Hurvitz, Rose S.

Drafting Room

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

Morss, Dorothy H.

Porter, Jean M.
Rollins, I. E.

Machine Shop

Keefe, J. B., Foreman
Barnet, F. J.
Bletzer, P. W.
Brennan, J.
Bunick, F. J.
Cabral, M., Jr.

Carter, C. E.
Gibbons, W. D.
Liljeholm, F. H.
MacDonald, J. R.
Marshall, J. J.
Muse, W. J.
Riemann, W.

Ryan, J. F.
Sanromá, J. B.
Shmid, E.
Smart, D. A.
Tucker, C. L.
Wentworth, W. G., Jr.

Secretaries

Austin, Carol M.
Barron, Gladys G.
Berg, Barbara G.
Bertozzi, Norma
Blais, Gisele G.
Boyajian, Judith A.
Brunetto, Deborah A.
Carbone, Angelina
Cavanaugh, Mary C.
Chapman, Carol A.
Cheever, Beatrice J.
Cohen, Phyllis J.
Collins, Brenda M.
Cummings, Jane F.
Daly, Marguerite A.
DiFranco, Frances D.

Dordoni, Joan M.
Epstein, Elinor F.
Geller, Elaine J.
Gordon, Linda S.
Greulach, Vicki E.
Imbornone, Elaine C.
Johnson, Barbara A.
Jordan, Kathleen
Kaloyanides, Venetia
Lannoy, Doris E.
Laurendeau, Carole A.
Loeb, Charlotte G.
Madden, Jean F.
May, Nancy A.
Mayman, Toby E.
McCarthy, Kathleen A.

McEntee, Doris C.
Milan, Marilyn A.
Morneault, Diane M.
Omansky, Betsey G.
Petone, Rosina C.
Sahagen, Judith A.
Scalleri, Mary B.
Smith, Claire F.
Solomon, Cynthia
Staffiere, Rose Carol
Thomson, Susan M.
Toebes, Rita K.
Touchette, Thelma B.
Townley, Madeline S.
Weisel, Linda E.
Vesey, Patricia A.

PERSONNEL

Technical Typists

Barnes, R. A.
Fleming, Patricia L.

Katan, Ann

Levine, R. I.
Rabkin, W. I.

Stock Rooms

Doiron, E. J., Foreman
Audette, A. G.
Cardia, P. F.
Haggerty, R. H.

Joyce, T. F.
Legier, D. O.
Lucas, W. G.
McDermott, J. F.

Pacitto, H. T.
Riley, J. F.
Sharib, G.
Sincuk, J., Jr.

Technician's Shop

Lorden, G. J., Foreman
D'Amico, C. R.

Fownes, Marilyn R.
Howell, W. B.

Lander, H. J.
MacDonald, K. B.

Tube Laboratory

Staff

Rosebury, F.
Ryan, L. W.

Glass Blowers

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

Technicians

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

PUBLICATIONS AND REPORTS

MEETING PAPERS PRESENTED

American College of Physicians Meeting, Cambridge, Massachusetts
March 4-8, 1963

    F. T. Hambrecht, Physical Methodology in Medical Research (invited)

Physics Department Colloquium, Columbia University, New York
March 15, 1963

    A. H. Barrett, Results of Mariner II Microwave Experiment (invited)

Medical College of Virginia, Richmond, Virginia
March 18, 1963

    P. A. Willis, Experimental and Analytical Dissection and Modeling of a Neurological Control System (invited)

Bionics Symposium 1963, Dayton, Ohio
March 19-21, 1963

    B. G. Farley, Aspects of the Behavior of a Neuron Network Model (invited)

    M. C. Goodall, Realizability of Inductive Logic

    K. Kornacher, How the Location of a Stimulus Is Correctly Signalled to the Central Nervous System

Symposium on Spectroscopy and Space Applications, California Institute of Technology, Pasadena, California
March 20-22, 1963

    A. H. Barrett, Results of Mariner II Microwave Experiment (invited)

Conference on College Composition and Communication, Los Angeles, California
March 23, 1963

    J. Viertel, Generative Grammars (invited)

Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
March 24, 1963

    G. L. Gerstein, Mathematical Models for Single Neuron Activity (invited)

Astronomy Department-Space Sciences Laboratory, University of California, Berkeley, California

March 25, 1963

    A. H. Barrett, Results of Mariner II Microwave Experiment (invited)

First Annual Symposium on Biomathematics and Computer Science in the Life Sciences, Houston, Texas

March 29, 1963

    W. A. Rosenblith, Some New Technologies and Their Promise for the Life Sciences (invited)

Northeastern New York Section Meeting, American Nuclear Society, New York

April 2, 1963

    D. J. Rose, Controlled Fusion Research at M. I. T. (invited)

Computation Laboratory, Harvard University, Cambridge, Massachusetts

April 3, 1963

    W. A. Rosenblith, Computers in Biology and Medicine (invited)

Symposium on "Feedback Systems Controlling Nervous Activity," Villahermosa, Tobasco, Mexico

April 8-10, 1963

    J. S. Barlow, Some Statistical Characteristics of Electrocortical Activity in Relation to Visual-Oculomotor Tracking in Man (invited)

    L. Stark, Principles of Neurological Feedback Control Systems (invited)

AIEE Automatic Control Symposium, Tufts University, Medford, Massachusetts

April 9, 1963

    P. A. Willis, Neuromuscular Control System of the Human Operator (invited)

Fourth Annual Symposium on Engineering Aspects of Magnetohydrodynamics, University of California, Berkeley, California

April 10-11, 1963

    R. J. Briggs and A. Bers, Electron-Beam Interactions with Ions in a Warm Plasma

Biennial Meeting, Society for Research in Child Development, University of California, Berkeley, California

April 11-13, 1963

    Paula Menyuk, Alternation of Rules in Children's Grammar

Thirty-fourth Annual Meeting, Eastern Psychological Association, New York
April 11-13, 1963

R. D. Hall, Some Characteristics of Cortical Evoked Potentials to Photic Stimuli
in the Behaving Rat

47th Annual Meeting, Federation of American Societies for Experimental Biology,
Atlantic City, New Jersey

April 16-20, 1963

A. Cavaggioni and M. Goldstein, Enhancement of Shock-Evoked Responses from
Striate Cortex Following either Onset or Termination of Light (Abstract to be
published in Federation Proceedings) (invited)

P. G. Katona and G. O. Barnett, An Analysis of Heart-Rate Control

American Geophysical Union, Washington, D.C.

April 17, 1963

A. H. Barrett, Results of Mariner II Microwave Experiment (invited)

113th Meeting, American Astronomical Society, Tucson, Arizona

April 17-20, 1963

D. H. Staelin, A. H. Barrett, and B. R. Kusse, Measurements of the Sun, Moon,
Venus, and Taurus A at 1.25-cm Wavelength

Graduate Seminar, Electrical Engineering Department, University of Rochester,
Rochester, New York

April 19, 1963

P. A. Willis, Modeling the Neuromuscular Control System of the Human Operator
(invited)

Symposium on Biophysics, American Physical Society, New York State Section,
Schenectady, New York

April 19, 1963

W. A. Rosenblith, Signal Processing in Biological Systems (invited)

American Physical Society Meeting, Washington, D.C.

April 22-25, 1963

G. Fiocco and E. Thompson, Techniques for Observing Thomson Scattering of
Optical Radiation from Electrons

C. W. Garland and J. S. Jones, Elastic Constants of Ammonium Chloride near
the Lambda Point

E. Thompson and G. Fiocco, Thomson Scattering of Optical Radiation from a
Thermal Plasma

Seminar on Biophysics for Science Writers, American Institute of Physics, Washington, D.C.

April 26, 1963

    W. A. Rosenblith, Biophysics of Communications Processes (invited)

Fourth National Symposium on Human Factors in Electronics, Men, Machines and Systems, Washington, D.C.

May 2, 1963

    L. Stark and E. Van Horn, A Computer Controlled Experiment in Human Prediction

Johnson Foundation Biophysical Seminar, University of Pennsylvania, Philadelphia, Pennsylvania

May 8, 1963

    L. Stark, Dynamical Approach to Biological Systems (invited)

Drexel Institute, Philadelphia, Pennsylvania

May 9, 1963

    L. Stark, Servoanalysis of Neurological Reflexes (invited)

National Academy of Sciences, Armed Forces-NRC Committee on Vision, Washington, D.C.

May 9, 1963

    L. Stark, Specialized Transducers and On-line Digital Computers for Setting Control Systems in the Eye (invited)

Sixty-fifth Meeting, Acoustical Society of America, New York

May 15-18, 1963

    D. M. Green, Consistency of Auditory Detection Judgments

    A. S. House and A. P. Paul, Automatic Reduction of Vowel Spectra: A Quantitative Analysis of Error Scores

    U. Ingard, Noise Generation of Fluid Flow in Ducts

Sixty-fifth Meeting, Acoustical Society of America, Symposium on Articulation in Speech Analysis and Synthesis, New York

May 15-18, 1963

    J. M. Heinz, Analysis of Speech Signals into Articulatory Parameters (invited)

Sixty-fifth Meeting, Acoustical Society of America, Symposium on Unit Activity in the Auditory Nerve, New York

May 15-18, 1963

N. Y-S. Kiang, Spontaneous Activity of Single Auditory Nerve Fibers in Cats (invited)

J. R. Welch, Psychoacoustical Study of Some Factors Affecting Human Echo Location (invited)

The Biological Bases of Memory, University of California, San Francisco Medical Center, San Francisco, California

May 21, 1963

W. A. Rosenblith, Sensory Information Processing, Neuro-electric Activity and Memory (invited)

1963 Spring Joint Computer Conference, American Federation of Information Processing Societies, Detroit, Michigan

May 22, 1963

W. A. Rosenblith, Computers and Brains: Competition and/or Co-existence (invited)

Lectures, Argonne National Laboratories, Lamont, Illinois

May 24, 1963

G. A. Brown, The Condensing Ejector Liquid-Metal MHD Power-Generation System (invited)

W. D. Jackson, Liquid-Metal MHD Generators for Space Power System (invited)

Conference on Computers in Medicine and Biology, New York Academy of Sciences — Section of Biology and Medicine and Division of Instruments, New York

May 27-29, 1963

W. E. Clark and C. E. Molnar, The LINC (invited)

L. Stark, P. A. Willis, A. A. Sandberg, S. Stanten, and J. Dickson, On-line Digital Computers Used in Biological Experiments and Modeling

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 39, Massachusetts.)

A. H. Barrett and E. Lilley, Mariner-2 Microwave Observations of Venus (Sky and Telescope)

G. Fiocco, Applicazioni dei Maser ottici alle ricerche spaziali (Missili)

JOURNAL ARTICLES ACCEPTED FOR PUBLICATION (continued)

G. L. Gerstein and B. Mandelbrot, Random-Walk Models for the Spike Activity of a Single Neuron (Biophys. J.)

R. J. Hull and H. H. Stroke, $^3S_1$ Hyperfine-Structure Separations and Isotope Shifts of Radioactive Cd^{107}, Cd^{109}, and of Stable Cd Isotopes (J. Opt. Soc. Am.)

U. Ingard and G. C. Maling, Jr., The Effect of Atmospheric Turbulence on Sound Propagated over Ground (J. Acoust. Soc. Am.)

G. C. Maling, Jr., Simplified Analysis of the Rijke Phenomenon (J. Acoust. Soc. Am.)

D. H. Staelin, A. H. Barrett, and B. R. Kusse, Measurement of the Sun, Moon, Venus, and Taurus A at 1.25-cm Wavelength (Astron. J.) (Abstract)

C. V. Stager, Hyperfine Structure of Hg^{197} and Hg^{199} (Phys. Rev.)

K. N. Stevens and A. S. House, Perturbation of Vowel Articulations by Consonantal Context: An Acoustical Study (J. Speech Hearing Res.)

J. A. Swets, Central Factors in Auditory Frequency Selectivity (Psychol. Bull.)

J. A. Swets and Susan T. Sewall, Invariance of Signal Detectability over Stages of Practice and Levels of Motivation (J. Exptl. Psychol.)

R. Weiss, Stark Effect and Hyperfine Structure of Hydrogen Fluoride (Phys. Rev.)

LETTERS TO THE EDITOR ACCEPTED FOR PUBLICATION

G. W. Stroke, Attainment of High Efficiencies in Blazed Optical Gratings by Avoiding Polarization in the Diffracted Light (Physics Letters (Amsterdam))

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 39, Massachusetts.)

411 Chung Laung Liu, Some Memory Aspects of Finite Automata

412 Sander Weinreb, A Digital Spectral Analysis Technique and Its Application to Radio Astronomy

SPECIAL PUBLICATIONS

S. C. Brown and G. Bekefi, Radiofrequency Emission from Plasmas not in Thermodynamic Equilibrium (Nuclear Fusion: 1962 Supplement, Part 3)

W. A. Clark and C. E. Molnar, The LINC (Proc. New York Academy of Sciences Conference on Computers in Medicine and Biology)

SPECIAL PUBLICATIONS (continued)

R. C. Gesteland, J. Y. Lettvin, W. H. Pitts, and A. Rojas, Odor Specificities of the Frog's Olfactory Receptors (Proc. First International Symposium on Olfaction and Taste. Pergamon Press, New York, 1963, pp. 19-34)

G. G. Hammes and P. Fassella, The Mechanism of Enzymatic Transamination (a chapter in Pyridoxal Catalysis. Pergamon Press, New York)

H. Putnam, An Examination of Grünbaum's Philosophy of Geometry (Philosophy of Science, Vol. II. Interscience Publishers, New York, 1963)

Introduction

This report, the seventieth in a series of quarterly reports issued by the Research Laboratory of Electronics, contains a review of the research activities of the Laboratory for the three-month period ending May 31, 1963. Since this is a report on work in progress, some of the results may not be final.

RADIO PHYSICS

I. PHYSICAL ELECTRONICS

Prof. W. B. Nottingham   R. F. DiGregorio   L. Schlessinger
J. L. Coggins   G. E. Johnson   L. E. Sprague

A. AN AUTOMATIC ELECTRONIC EMISSION CONTROL

In association with a number of electronic devices, including ionization gauges, mass spectrometers, and others, it is desirable to have a means by which the electron emission delivered to a collector can be stabilized and maintained with the minimum of attention at any preassigned value. Specifically, as it relates to the ionization gauge, it is desirable to have a control that will maintain an electron emission current accurately within 1 per cent at any value from 1 μa to 100 ma. Although rather complex, a circuit[1] has been developed which nearly meets these characteristics. This circuit depends very largely on a multiplicity of vacuum-tube devices. It operates the emitter filament on rectified direct current and controls the emission at any preassigned value within a few per cent. A parallel development that depends almost entirely on solid-state devices, and includes only one vacuum tube is nearly complete. This tube may be replaced later by a "field-effect" transistor. Preliminary tests on two of the ranges, 0-10 μa and 0-100 μa, yielded preassigned currents with an accuracy much better than 1 per cent.

Another closely related development pertains to a transistorized dc amplifier that has been designed to furnish the coupling between a shunt resistance and an output meter that will then indicate accurately full-scale responses from 1 μa, or more, to indefinitely large values that depend only on the shunt resistance. Although sensitive multimeters are available in the laboratory with full-scale sensitivity from 10 μa to larger values, they are easily damaged by severe accidental overloads. The circuit now being developed will permit overloads of 100 to 500 times the normal full scale without damage to either the electronic coupling or the indicating meter.

Circuit details and specifications of these developments will be made available as soon as they have been thoroughly tested.

W. B. Nottingham

References

1. G. E. Johnson, Design and Construction of an Electronic Emission Regulator, S. B. Thesis, Department of Physics, M. I. T., June 1963.

B. ENERGY LEVELS OF THE CESIUM ATOM

In order to understand the details of operation of a thermionic energy converter that depends on the ionization of cesium in the space between the electrodes, it is desirable

to have a listing of the electronic energy levels. These levels have been evaluated, and are listed along with their energy-level identifications. All values are based on the analysis given by Grotrian.[1] Not only are the transitions expressed in terms of their

Table I-1. Principal Series.

| Terms | | λ | Electron-Volts from | Electron-Volts from |
Final	Initial	(Å)	Ground State	Ionization Potential
$6\,^2S_{1/2}$	$6\,^2P_{1/2}$	8943.46	1.386	2.507
$6\,^2S_{1/2}$	$6\,^2P_{3/2}$	8521.12	1.455	2.438
$6\,^2S_{1/2}$	$7\,^2P_{1/2}$	4593.16	2.699	1.194
$6\,^2S_{1/2}$	$7\,^2P_{3/2}$	4555.26	2.721	1.172
$6\,^2S_{1/2}$	$8\,^2P_{1/2}$	3888.65	3.188	0.705
$6\,^2S_{1/2}$	$8\,^2P_{3/2}$	3876.39	3.198	0.695
$6\,^2S_{1/2}$	$9\,^2P_{1/2}$	3617.41	3.427	0.466
$6\,^2S_{1/2}$	$9\,^2P_{3/2}$	3611.52	3.433	0.460
$6\,^2S_{1/2}$	$10\,^2P_{1/2}$	3480.13	3.562	0.331
$6\,^2S_{1/2}$	$10\,^2P_{3/2}$	3476.88	3.566	0.327
$6\,^2S_{1/2}$	$\infty\,^2P$	3184.2	3.893	0

corresponding wavelengths, but these wavelengths have been evaluated individually, and the corresponding energy states expressed in electron-volts relative both to the ground state and the ionization potential are given. These results are recorded in Tables I-1 through I-4.

Table I-2. Sharp Series.

| Terms | | λ | Electron-Volt | Electron-Volts | Electron-Volts |
Final	Initial	(Å)	Difference in Levels	from Ground State	from Ionization Potential
$6\,^2P_{3/2}$	$7\,^2S_{1/2}$	14694.8	0.8436	2.299	1.594
$6\,^2P_{1/2}$	$7\,^2S_{1/2}$	13588.1	0.9123		
$6\,^2P_{3/2}$	$8\,^2S_{1/2}$	7944.11	1.561	3.016	0.877
$6\,^2P_{1/2}$	$8\,^2S_{1/2}$	7609.13	1.629		
$6\,^2P_{3/2}$	$9\,^2S_{1/2}$	6586.94	1.882	3.337	0.556
$6\,^2P_{1/2}$	$9\,^2S_{1/2}$	6354.98	1.951		
$6\,^2P_{3/2}$	$10\,^2S_{1/2}$	6034.6	2.054	3.509	0.384
$6\,^2P_{1/2}$	$10\,^2S_{1/2}$	5839.11	2.123		
$6\,^2P_{3/2}$	$\infty\,^2S_{1/2}$	5085	2.438	3.893	0
$6\,^2P_{1/2}$	$\infty\,^2S_{1/2}$	4945	2.507		0

Table I-3. Diffuse Series.

| Terms | | λ | Electron-Volt | Electron-Volts | Electron-Volts from |
Final	Initial	(Å)	Difference in Levels	from Ground State	Ionization Potential
$6\,^2P_{3/2}$	$5\,^2D_{3/2}$	36127	0.3432	1.798	2.095
$6\,^2P_{3/2}$	$5\,^2D_{5/2}$	34892	0.3553	1.810	2.083
$6\,^2P_{1/2}$	$5\,^2D_{3/2}$	30100	0.4119		
$6\,^2P_{3/2}$	$6\,^2D_{3/2}$	9208.4	1.346	2.801	1.092
$6\,^2P_{3/2}$	$6\,^2D_{5/2}$	9172.23	1.352		
$6\,^2P_{1/2}$	$6\,^2D_{3/2}$	8761.35	1.415	2.807	1.086
$6\,^2P_{3/2}$	$7\,^2D_{3/2}$	6983.37	1.775	3.230	0.663
$6\,^2P_{3/2}$	$7\,^2D_{5/2}$	6973.17	1.778	3.233	0.660
$6\,^2P_{1/2}$	$7\,^2D_{3/2}$	6723.18	1.844		
$6\,^2P_{3/2}$	$8\,^2D_{3/2}$	6217.27	1.994	3.449	0.454
$6\,^2P_{3/2}$	$8\,^2D_{5/2}$	6212.87	1.995	3.450	0.453
$6\,^2P_{1/2}$	$8\,^2D_{3/2}$	6010.33	2.063		
$6\,^2P_{3/2}$	$\infty\,^2D_{3/2}$	5085	2.438	3.893	0
$6\,^2P_{1/2}$	$\infty\,^2D_{1/2}$	4945	2.507	3.893	0

Table I-4. Fundamental Series.

| Terms | | λ | Electron-Volt | Electron-Volts | Electron-Volts |
Final	Initial	(Å)	Difference in Levels	from Ground State	from Ionization Potential
$6\,^2D_{5/2}$	$4\,^2F_{7/2}$	10124.7	1.224	3.034	0.859
$6\,^2D_{5/2}$	$4\,^2F_{5/2}$	10120.0	1.225		
$6\,^2D_{3/2}$	$4\,^2F_{5/2}$	10025.1	1.237		
$6\,^2D_{5/2}$	$5\,^2F_{7/2}$	8079.02	1.534	3.344	0.548
$6\,^2D_{5/2}$	$5\,^2F_{5/2}$	8078.92	1.534		
$6\,^2D_{3/2}$	$5\,^2F_{5/2}$	8015.71	1.547		
$6\,^2D_{5/2}$	$6\,^2F_{7/2}$	7279.95	1.703	3.513	0.380
$6\,^2D_{5/2}$	$6\,^2F_{5/2}$	7279.89	1.703		
$6\,^2D_{3/2}$	$6\,^2F_{5/2}$	7228.53	1.715		
$6\,^2D_{5/2}$	$\infty\,^2F$	5952	2.083	3.893	0
$6\,^2D_{3/2}$	$\infty\,^2F$	5917	2.095	3.893	0

W. B. Nottingham

References

1. W. Grotrian, Graphische Darstellung der Spektren von Atomen und Ionen mit ein, zwei und drei Valenzelektronen, Struktur der Materie in Einzeldarstellungen VII, M. Born and J. Franck (eds.) (Springer Verlag, Berlin, 1928), Bd. 2, p. 40.

II. MICROWAVE SPECTROSCOPY[*]

Prof. M. W. P. Strandberg
Prof. R. L. Kyhl
Dr. B. D. Nageswara Rao
J. M. Andrews, Jr.
Y. H. Chu
R. Huibonhoa

J. G. Ingersoll
P. F. Kellen
J. D. Kierstead
S. H. Lerman
J. W. Mayo

H. Pauwels
Mahin Rahmani
W. J. Schwabe
G. J. Sehn
J. R. Shane
C. F. Tomes

A. WORK COMPLETED

1. SPIN-LATTICE RELAXATION IN PARAMAGNETIC CRYSTALS

This work has been completed by J. R. Shane and submitted as a thesis to the Department of Physics, M. I. T., May 10, 1963, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. An abstract of the thesis follows.

A detailed investigation of the spin-phonon interaction for iron group spins at liquid-helium temperatures is presented. An effective interaction Hamiltonian for single phonon processes is used to describe the lattice-induced transitions between pairs of spin levels and the recovery of the energy level populations to thermal equilibrium.

The rate equations for spin-lattice relaxation are shown to follow directly from a density matrix formulation of the problem. The usual time-dependent perturbation theory expressions for the transition probabilities appear naturally in this formalism. Solutions of these equations are obtained for special magnetic field orientations for which direct relationships exist between the transition probabilities and the observed relaxation times.

A Green's function method is developed for calculating the matrix elements of the lattice operators that contribute to the interaction. This approach avoids the usual assumption that the normal lattice modes are unaffected by the presence of the paramagnetic impurities. The theory is applied to the phonon-impurity coupling in ruby, and the results are compared with the perfect lattice approximation. It is shown that the coupling is controlled by the defect states of the lattice, and these effects must be considered in order to obtain spin-phonon transition probabilities that have the observed frequency dependence.

Relaxation experiments at the symmetric orientation in ruby are described and the coefficients of the rank-two spin tensors that appear in the predicted form of the interaction are extracted from the data. An empirical interaction Hamiltonian is constructed with these coefficients and the frequency dependence that results from the Green's function method. This form of the interaction is used to calculate relaxation times at the

---

90° orientation for several magnetic field values and temperatures. Good quantitative agreement with published relaxation measurements has been found.

2. NONLINEAR EFFECT OF LASER RADIATION ON METALS

This work has been completed by Y. H. Chu and submitted as a thesis to the Department of Physics, M. I. T., May 1963, in partial fulfillment of the requirements for the degree of Bachelor of Science. An abstract of the thesis follows.

An attempt was made to observe the optical nonlinear effect from the interaction of the ruby laser radiation and the surface of metals, in particular, the generation of second harmonics of the laser frequency by the metal. By using a crude free-electron theory, this phenomenon had a minimum theoretical effective cross section of $2 \times 10^{-32} cm^2$. The electric fields that are necessary to allow the second-order components of the induced dipole are much greater than those in insulators and semiconductors. After 30 experiments in which samples of copper and aluminum were used, no positive effect was observed. Detection was accomplished by photographic emulsion, which had an experimental minimum detectability of 1,000 photons.

3. THE OPERATION AND MODE SPECTRUM OF A CONFOCAL RESONANCE CAVITY AT 23,870 mc WITH A CONSIDERATION OF THE AMMONIA ABSORPTION COEFFICIENT

This work has been completed by C. F. Tomes and submitted to the Department of Physics, M. I. T., May 1963, as a thesis in partial fulfillment of the requirements for the degree of Bachelor of Science. An abstract of the thesis follows.

A confocal resonator with spherical ends (b=16.2 cm) was constructed and operated in the K-band microwave region around 23,870 mc. The nondegenerate mode spectrum was fully explained by having different eigenfunctions for the m and n valves. In removing this last degeneracy in m and n, the split of 7 mc (independent of separation) was found to correspond to a difference of 4 mm in radii of curvature. Near the confocal separation a high-loss region was seen to be just as predicted by these different curvatures. Because of technical problems, determination of the ammonia absorption coefficient was not possible.

M. W. P. Strandberg

B. ULTRASONIC ATTENUATION IN SUPERCONDUCTORS

Measurements of ultrasonic attenuation in superconducting metals have been extended to the L-band region. The experiment[1] is identical to that carried out at 165 mc except for certain changes in the equipment. The transmitter is a modified General Radio Unit Oscillator. The dc plate supply lead has been disconnected and 1200-volt dc pulses of
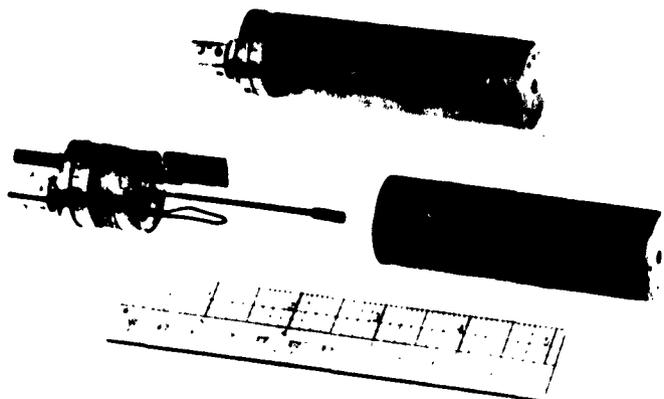
Fig. II-1.

Re-entrant cavities used for the L-band transmitting and receiving transducers.

0.5-μsec duration obtained from an external source are applied to the plate of the oscillator tube. The oscillator provides approximately 25 watts of peak power. The re-entrant cavities are shown in Fig. II-1. Coarse tuning is achieved by rotating the barrel; this operation changes the gap length between the end of the post and the end wall of the cavity in which the quartz rod is inserted. Also, a fine-tuning control is included in one of the cavities, which consists of a small vane that can be rotated through 180° with respect to the rf magnetic field within the cavity. This provides a means of matching the resonant frequencies of the cavities while they are at helium temperature. In all other respects the details of the experiments are similar to those reported earlier.[1]

Interaction of 1 kmc coherent phonons with incoherent thermal phonons at room temperature is quite pronounced. The result is that with our equipment it is never possible to see more than five or six acoustic echoes in 1/2-inch quartz rods at room temperature; however, the number of observable echoes at 4.2°K is around 100. The attenuation coefficient in quartz, a_q, was measured at both temperatures at 0.89 kmc: at 300°K, $a_q = 4.3$ db/cm; at 4.2°K, $a_q = 0.98$ db/cm. The phonon-phonon scattering is negligible at 4.2°K, so that the 0.98 db/cm was considered to be the residual temperature-independent loss resulting from lattice defects and geometrical effects. This residual attenuation was subtracted from both measurements, and the resulting 3.3 db/cm was compared with, and found to be consistent with, the temperature-dependent phonon-phonon scattering reported by Jacobsen, Bömmel, and Dransfeld.[2]

Attention was then turned to the measurement of the difference in the electronic contribution to the ultrasonic attenuation coefficient between normal and superconducting indium. Three samples of indium of varying thickness were sandwiched between quartz rods, and the height of the transmitted acoustic pulses was measured as a function of external magnetic field.

Sample 1   Thickness: 0.015 cm ± 0.001

                                      $T = 1.9°K ± 0.1°K$

The transmitted pulses are shown in Fig. II-2a for the normal state and in Fig. II-2b for the superconducting state. The pulse denoted a_0 is rf leakage from the transmitter. The pulse denoted a_1 is the first acoustic pulse that has traversed the sample. The pulse denoted a_2 is a superposition of two acoustic pulses: one has made a double reflection in the first rod; the other, a double reflection in the second rod. The superposition is caused by the fact that the two quartz transducer rods are equal in length (within ±.001 in.). The pulse denoted Δ is an rf calibration pulse controlled by a calibrated attenuator and variable delay, which has been introduced into the system for the purpose of measuring the height of the acoustic pulses. The pulse denoted a_3 is a superposition of three acoustic pulses: one has made four reflections in the first rod, another has made a double reflection in the first rod and a double reflection in the second rod, and the third has made four reflections in the second rod. Since a_3 was the largest pulse, measurements of the magnetic field dependence of the ultrasonic attenuation coefficient were made on it alone. These results are plotted in Fig. II-3. The difference in magnitudes of the other pulses are also shown for comparison.

As mentioned in an earlier report[1] $(\alpha_n - \alpha_s)$ can probably not be measured from any pulses other than the first because of the unknown phase relationships between the various pulses forming the superpositions. The exact cause of the phase changes between these pulses as the sample is switched between the superconducting and the normal states is not understood, but the existence of such an effect seems to be the best explanation for the discrepancies in $(\alpha_n - \alpha_s)$ shown in Fig. II-3.

The magnetic field dependence of $(\alpha_n - \alpha_s)$ for this sample supports the contention stated previously[1] that the shape of the curve is characteristic of the intermediate state. Note the very sharp transition at $(H_c)_{exp}$. This indicates that there is virtually no region of the magnetic field where this sample is in the intermediate state. As previously given[3] $H_e/H_c = 1 - n$. Our data indicate that $H_e \approx H_c$, and we conclude that $n = 0$ for Sample 1. For any given body the sum of the demagnetizing coefficients ($4\pi n$) for any three mutually perpendicular axes is 4π. Since $n=1$ for a thin plane disc with its surface normal to the magnetic field, $n = 0$ for a disc with its surface parallel to the magnetic field. The ratio of the diameter to the thickness of Sample 1 is 20, and therefore the agreement with the macroscopic theory of the intermediate state is excellent. The values $[H_c(T)]_{calc}$ shown in Figs. II-3, II-5, and II-7 were obtained from the well-known deviations of superconducting critical fields from the parabolic law.

Measurements of α_n in a sample as thin as this are likely to be unreliable for two reasons. First, the magnitude of the change in pulse height is so small that it introduces appreciable error. (Compare a_2 in Fig. II-2a and 2b.) The second is that it is quite
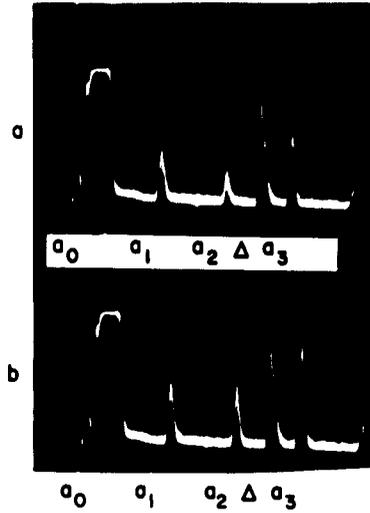
Fig. II-2. (a) Ultrasonic pulses at L-band transmitted through Sample 1 of indium
in the normal state. a_0 is rf leakage; a_1, the first transmitted pulse;
a_2, a superposition of two reflected pulses; Δ, a calibration pulse; a_3,
a superposition of three reflected pulses (scale, 2 μsec/cm).
(b) Same as in (a) except that the sample is in the superconducting state.
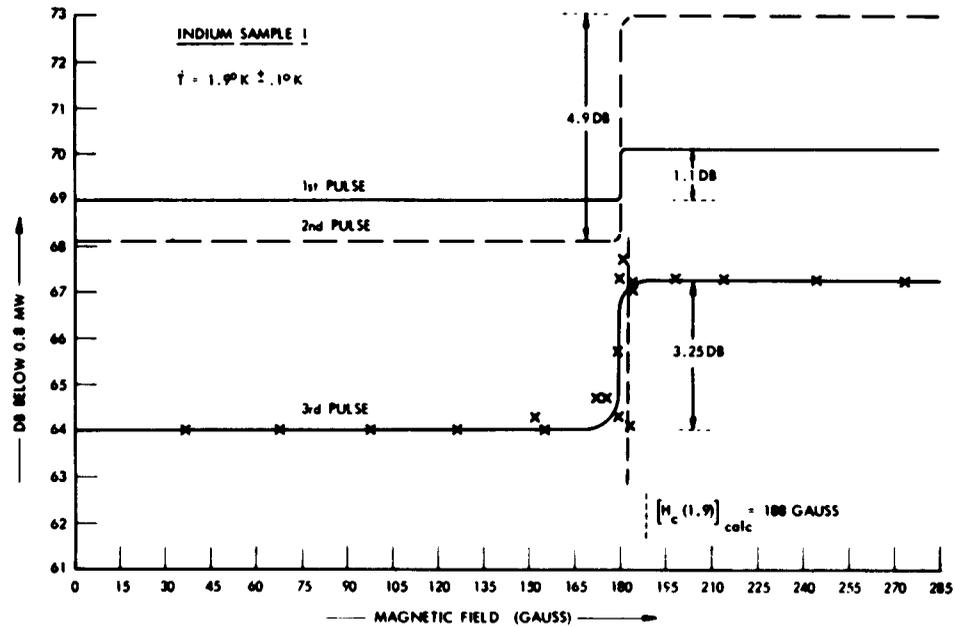
Fig. II-3. Attenuation of L-band ultrasonic pulses in Sample 1 of indium plotted as
a function of external magnetic field. Experimental points were taken only
on the third pulse.

possible that multiple reflections are taking place in such a thin sample. Multiple reflec-
tions have never been observed in samples thick enough for them to be resolved. On the
other hand, Sample 1 is so thin (1/18 of the pulse length) that up to 10 reflections could
take place without noticeable pulse broadening. However, setting these objections aside
for the moment, we tentatively calculate a_n for Sample 1, using the formula given pre-
viously.[4]

$$(a_n)_{exp} \approx 17 \text{ cm}^{-1} \quad \text{(tentative value).}$$

From the Pippard theory for ultrasonic attenuation in metals at low temperatures a',
evaluated at 910 mc, is

$$(a')_{th} = 23 \text{ cm}^{-1} \quad q\ell \gg 1,$$
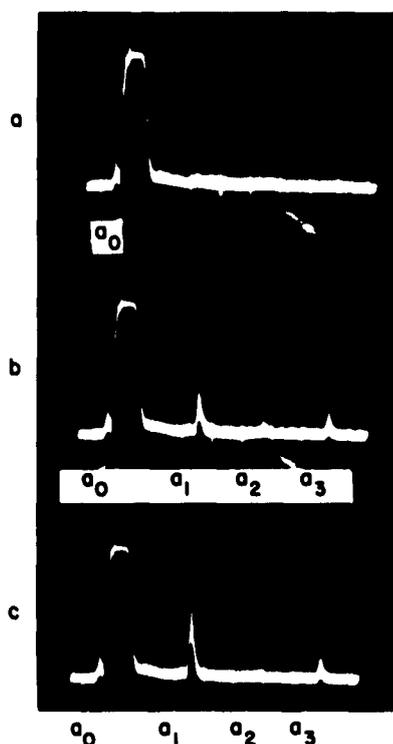
Fig. II-4. (a) Ultrasonic attenuation at L-band in Sample 2 of indium in the normal
state. No observable acoustic energy traverses the sample.
(b) Same as (a), but in the intermediate state. Note the pulse a_2, made
up of the superposition of the two double reflections. The appearance
of this pulse was a random function of the manner in which the magnetic
field was changed.
(c) Same as (a), but in the superconducting state. Note that a_2 has almost
entirely disappeared.

where a' denotes the limiting value of the electronic contribution to the ultrasonic attenuation coefficient for long electronic mean-free path.

Note that in all of the measurements on indium reported here, the temperature of the sample was sufficiently low that the superconducting energy gap had opened up to a value very close to its value at absolute zero. That is, the energy-gap factor[4] $\tanh\left[\dfrac{\epsilon(T)}{2kT}\right]$ was very nearly equal to unity. This means that when the sample is switched from the superconducting to the normal state we observe very nearly the full effect on the ultrasonic attenuation that is due to the conduction electrons.

Sample 2   Thickness: 0.208 cm ± 0.001

                            T = 1.8°K ± 0.1°K

Figure II-4 shows the acoustic pulses transmitted through this sample. Figure II-4a shows that in the normal state no transmitted sound can be observed above the noise.
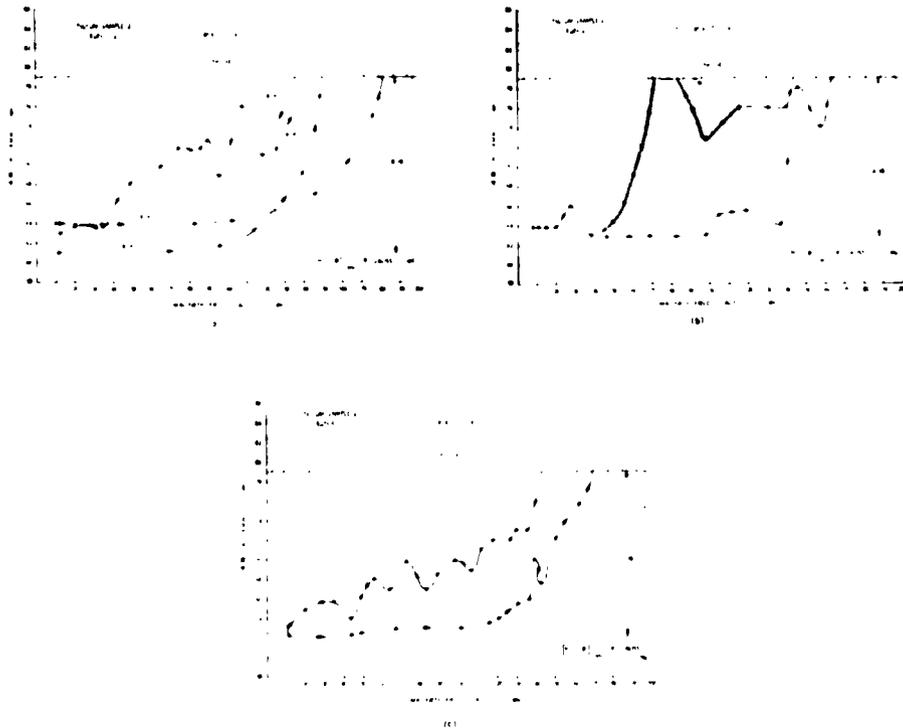
Fig. II-5. Attenuation of L-band ultrasonic pulses in Sample 2 of indium plotted as a function of external magnetic field. The shaded area shown in (b) indicates a region of extreme instability where the pulse height fluctuated even while the magnetic field was held constant.

In Fig. II-4b the sample is in the intermediate state; in Fig. II-4c, the superconducting state. The pulse denoted by a_2 represents just such a superposition of pulses as discussed with respect to Sample 1. Note that this pulse appears here only in the intermediate state. Its appearance was not reproducible, sometimes appearing momentarily and then disappearing altogether. At other times, it could be displayed for long periods of time. Its magnitude seemed to depend to some extent on the speed at which the magnetic field was varied and on other details of the traversal through the intermediate state. The behavior of this pulse seemed to indicate some sort of instability that characterized the sample while in the intermediate state. It is thought that the height of this pulse may be determined by the phase relationship between the individual pulses composing the superposition. These phase differences, in turn, may be caused by the structure of the intermediate state which allows acoustic energy to pass through different sections of the sample, the thickness of the sample possibly varying by one or more acoustic wavelengths over the diameter of the quartz transducer rods.

Such instability was even more evident in the magnetic field dependence of the ultrasonic attenuation and can be seen in the plots shown in Fig. II-5. Note that the actual structure of the curve seems to be quite random. In some cases, especially during run 3, a good deal of instability in the pulse height could be observed when the magnetic field was not being altered at all. The instability effects, however, were usually sufficiently well correlated with magnetic field changes to rule out instrumentation difficulties. The shaded portion of the curve in Fig. II-5b represents a region of extreme instability. Each run, however, showed a reproducible hysteresis effect. Similar effects have been observed by other investigators in polycrystalline tin and lead.[6,7] This behavior has been attributed to the trapping of magnetic flux in the center of the sample which causes a persistence of the intermediate state as the magnetic field decreases.

We should point out one feature of these observations. We are working with such a combination of high frequency and thick sample that the total change in attenuation $(a_n - a_s)$ between the normal and the superconducting state is very large. Thus when the sample is in the normal state, the fraction of the acoustic power reaching the receiver is of the order of one one-hundredth that which is passed while the sample is in the superconducting state. The obvious disadvantage of these conditions is that the total change in attenuation could not be measured in this sample. However, one distinct advantage is that small fluctuations of the ultrasonic attenuation coefficient of the metal are greatly magnified in the intermediate state. Thus it is possible that such techniques may provide detailed information concerning the kinetics of the superconducting transition process as the metal passes through the intermediate state.

Although this sample was too thick for us to measure the total change in the ultrasonic attenuation coefficient between the normal and the superconducting states, we can conclude that it is somewhat greater than 18 db. Since the pulse entered the noise close

to the critical field, most of the change may have been observed. Again, we tentatively calculate the value of the attenuation coefficient a_n for Sample 2.[4]

$$(a_n)_{exp} \gtrless 20 \text{ cm}^{-1} \qquad \text{(estimated value).}$$

Sample 3   Thickness: 0.127 cm ± 0.001

                                     T = 2.0°K and 1.8°K ± 0.1°K

Figure II-6a, 6b, and 6c shows the pulses transmitted through Sample 3 in the superconducting, intermediate, and normal states, respectively. They are denoted a_1, a_2, and a_3: a_0 is rf leakage from the transmitter. The origin of the tiny pulse denoted b (just before a_1) is unknown. Its position indicates that if it is an acoustic pulse that passed through the sample, its velocity must have been much greater than the usual velocity of
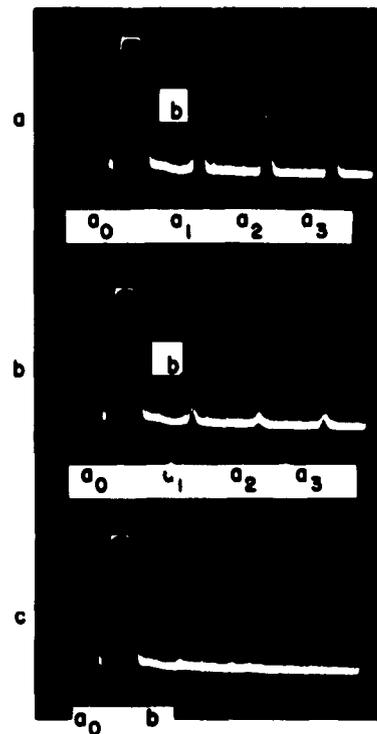
Fig. II-6.   (a) Ultrasonic pulses at L-band transmitted through Sample 3 of indium in the superconducting state.
                 (b) Same as (a), but in the intermediate state.
                 (c) Same as (a), but in the normal state. The a's have disappeared entirely. Note the tiny pulse, denoted b. Its origin is unknown, but if it is an acoustic pulse it has a velocity much greater than that of the longitudinal mode denoted by the a's.
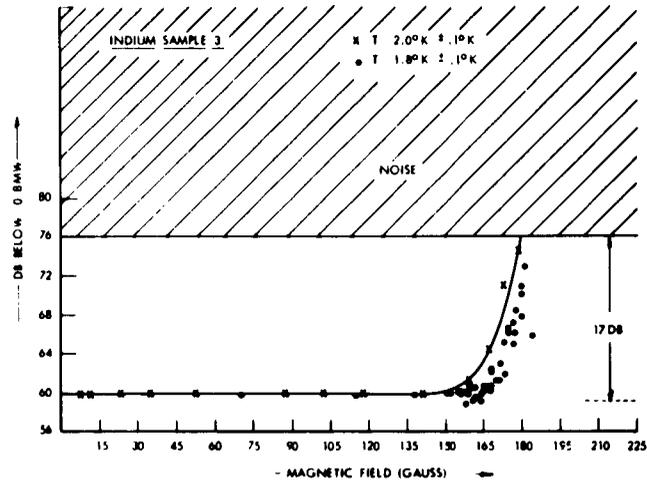
(a)
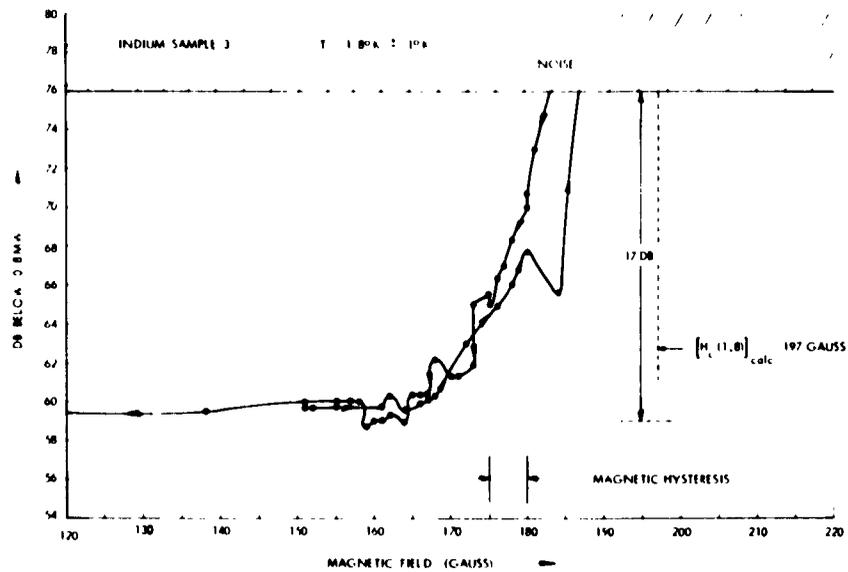
(b)

Fig. II-7. (a) Attenuation of L-band ultrasonic pulses in Sample 3 of indium plotted as a function of external magnetic field. Note the absence of the superconducting hysteresis effect.

(b) Expanded view of the points for T = 1.8°K, plotted in (a), showing the details of the magnetic path through the intermediate state. If any rconducting hysteresis exists it would be masked by the hysteresis of the external magnet.

longitudinal ultrasonic waves in indium, which is 2.72 km/sec along the 100 axis at 4.2°K.[8] It is not known whether our samples were single crystals; however, the orientation dependence of the velocity of the longitudinal mode is small,[8] and our samples were so thin that accurate velocity determinations were not possible.

The data taken on the magnetic field dependence of the ultrasonic attenuation coefficient show no large hysteresis effects. This is evident in Fig. II-7. The spread in the data occurs throughout a region that is so narrow that much of it should probably be attributed to hysteresis in the iron of the external electromagnet, which may have been as much as 5 gauss. However, the magnitude of this uncertainty was sufficiently small that no attempt was made to correct for magnetic hysteresis in any of the data. Again, the bond was too thick to obtain a determination of the entire change in the attenuation $(a_n - a_s)$, but data were taken up to a magnetic field of approximately 182 gauss, which was the field in which the sharp superconducting transition occurred for Sample 1. Consequently, this is probably the most reliable data taken thus far from which a_n at 910 mc can be estimated. We obtain[4]

$$(a_n)_{exp} = 28 \text{ cm}^{-1} \quad \text{(estimated value)}.$$

The most puzzling aspect of the data taken for Samples 2 and 3 is the pronounced difference in the superconducting hysteresis effects. The samples were prepared from the same source, and their exterior dimensions did not differ greatly. Not much information, however, is known about the details of the acoustic path through the sample. That these effects are important in the observation of superconducting hysteresis effects has been demonstrated by Mackinnon.[6]

The author is indebted to John Kierstead for technical assistance with the pulse-forming circuit, and to E. C. Ingraham for the construction of the L-band cavities.

J. M. Andrews, Jr.

References

1. J. M. Andrews, Jr., Ultrasonic attenuation in superconductors, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M.I.T., April 15, 1963, pp. 1-9.

2. E. H. Jacobsen, Experiments with phonons at microwave frequencies, Quantum Electronics, edited by C. H. Townes (Columbia University Press, New York, 1960), p. 468.

3. J. M. Andrews, Jr., op. cit., p. 8, Eq. 7.

4. Ibid., p. 7, Eq. 5.

5. A. B. Pippard, Phil. Mag. 46, 1104 (1955).

6. L. Mackinnon, Phys. Rev. 106, 70 (1957).

7. H. E. Bömmel, Phys. Rev. 96, 220 (1954).

8. B. S. Chandrasekhar and J. A. Rayne, Phys. Rev. 124, 1011 (1961).

III. FAR INFRARED SPECTROSCOPY

Prof. C. H. Perry

A. TECHNIQUES OF FAR INFRARED STUDIES

1. Introduction

The properties of solids in the far infrared part (50-1000 microns) of the spectrum have hardly been explored. In this region lie the resonances of many crystal lattices and excitations of electrons in superconductors, and we may expect to find here other phenomena that are characteristic of collective oscillations in solids, such as resonances of antiferromagnetics and other coupled-spin systems.

The reasons for the deficiencies of spectroscopic technique in this region are well known. The methods of microwave spectroscopy, in which powerful coherent sources are used, of course, have very high sensitivity. There is great difficulty, however, in the progressive extension of the technique to shorter wavelengths. Although sources of 1-mm radiation have been available for some time, the only work thus far reported at wavelengths less than 1 mm has been the accurate determination of frequencies of rotation of certain molecules. This technique employs harmonics derived from sources of radiation in the millimeter range, and a grating can be used with this radiation to sort out the various harmonics, provided that there is enough overlap that an unknown spectrum can be determined over a wide range of frequencies.

The techniques of grating spectrometry with the use of hot untuned sources have been extended beyond 150 μ by various workers.[1-7] Sensitive detectors are essential because the energy from a hot source falls off rapidly with increasing wavelength (only one part in 10^8 of the total emission of a black body at 10,000°K lying between, say, 150 μ and 350 μ). Furthermore, the energy in the various orders reflected from an echelette grating in a given direction varies as the cube of the order number for unfiltered incident black-body radiation, so that extremely effective filtering is necessary if the intrusion of high-order spectra is to be avoided. In this report improved techniques are described for studying the spectral properties of solids from 373°K down to 4°K, of liquids from 373°K down to 77°K, and the absorption spectra of gases in cells with up to 1-meter path length.

We shall continue investigation of sources, detectors, and filters in order to obtain the maximum performance from the spectrometer. The results thus far obtained will be discussed in the following sections.

Current solid-state research problems include the infrared spectra of perovskite titanates, antiferromagnetic materials, solids exhibiting ferroelectricity, and inorganic compounds that have low internal molecular and lattice vibrations. We propose to investigate materials that show free-carrier absorption and may have electronic transitions in the far infrared at low temperatures (for example, band transitions in semiconductors,

and the absorption spectra of some rare earth salts).

2. Design and Operation of the Spectrometer

The design and construction of this spectrometer has been carried out in collaboration with the Spectroscopy Laboratory, M. I. T. A complete remodification of the original long-wave instrument constructed by Lord and McCubbin[4] has been completed.

The optical design of the spectrometer is illustrated in the ray diagram of Fig. III-1, in which the monochromator unit is shown in conjunction with the low-temperature bolometer. Figure III-2 is a photograph of the far infrared spectrometer. The flexibility of the sample area, A, is illustrated. A modified Perkin-Elmer reflection attachment for reflection measurements at nearly normal incidence is shown intercepting the beam. This can easily be moved and replaced with other solid, liquid or gas sample holders. The kinematically mounted mirror, K, is removable so that a conventional Golay detector, D, can be used for preliminary measurements and low-resolution studies.

The entire spectrometer can be evacuated to remove water vapor that absorbs very strongly in the far infrared. A rough vacuum of 10^{-1} mm Hg is quite adequate for this purpose. The walls, V, are demountable for major adjustments to the optical system, and the slit mechanism, L, and the grating drive, G, are brought out through "O"-ring seals for maximum convenience of operation.

The source, S, is a medium-pressure mercury discharge lamp (General Electric H-4) in a quartz envelope, forming a 1:1 image at position N. A Nernst glower can be located in this position for operating the spectrometer at shorter wavelengths. Rotating crystal choppers of NaCl, KBr, and CsI are used near the focus at position C. They modulate the long-wavelength radiation at 13 cps, are transparent to the shorter wavelengths, and have a relatively sharp change in transmission. A number of radial wires are located in the open section of the chopper to compensate for the dielectric reflection of the crystal, and the diameter of the wires is adjusted to minimize the modulation at short wavelengths. The mirror at position P is kinematically mounted, and can be replaced by a scatter plate, reststrahlen filter or filter grating. A 2:1 magnified image of the source is produced at the entrance slit, W_1.

The Perkin-Elmer Model 98 monochromator has the grating, G, in the Littrow mounting. An off-axis paraboloid, O, and the entrance and exit slits, W_1 and W_2, open up to a width of 10 mm. A series of Bausch and Lomb gratings (64 mm × 64 mm area) is used as dispersing elements, and the gratings are blazed at 444, 333, 222, 89, 44.5, and 22 cm^{-1}; the 89 cm^{-1} grating is used in both the first and second orders. For use with the Golay detector the three holders, Q_1, Q_2, and Q_3, contain reststrahlen filters BaF_2, NaCl, KCl, KBr, CsBr, TlBr or KRS-5, the choice depending on the region to be investigated. Normally, only two filters are required; the third is often used for
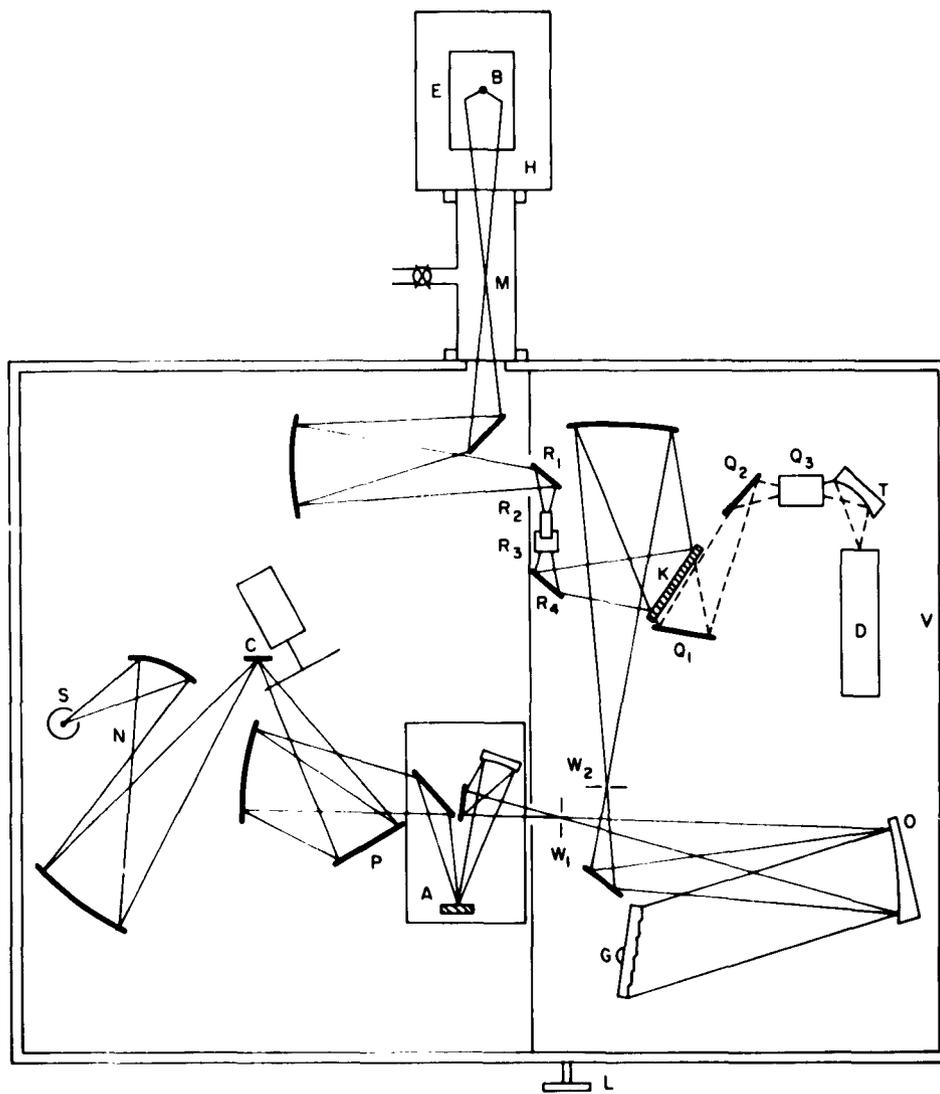
Fig. III-1.   Optical layout of the far infrared spectrometer.
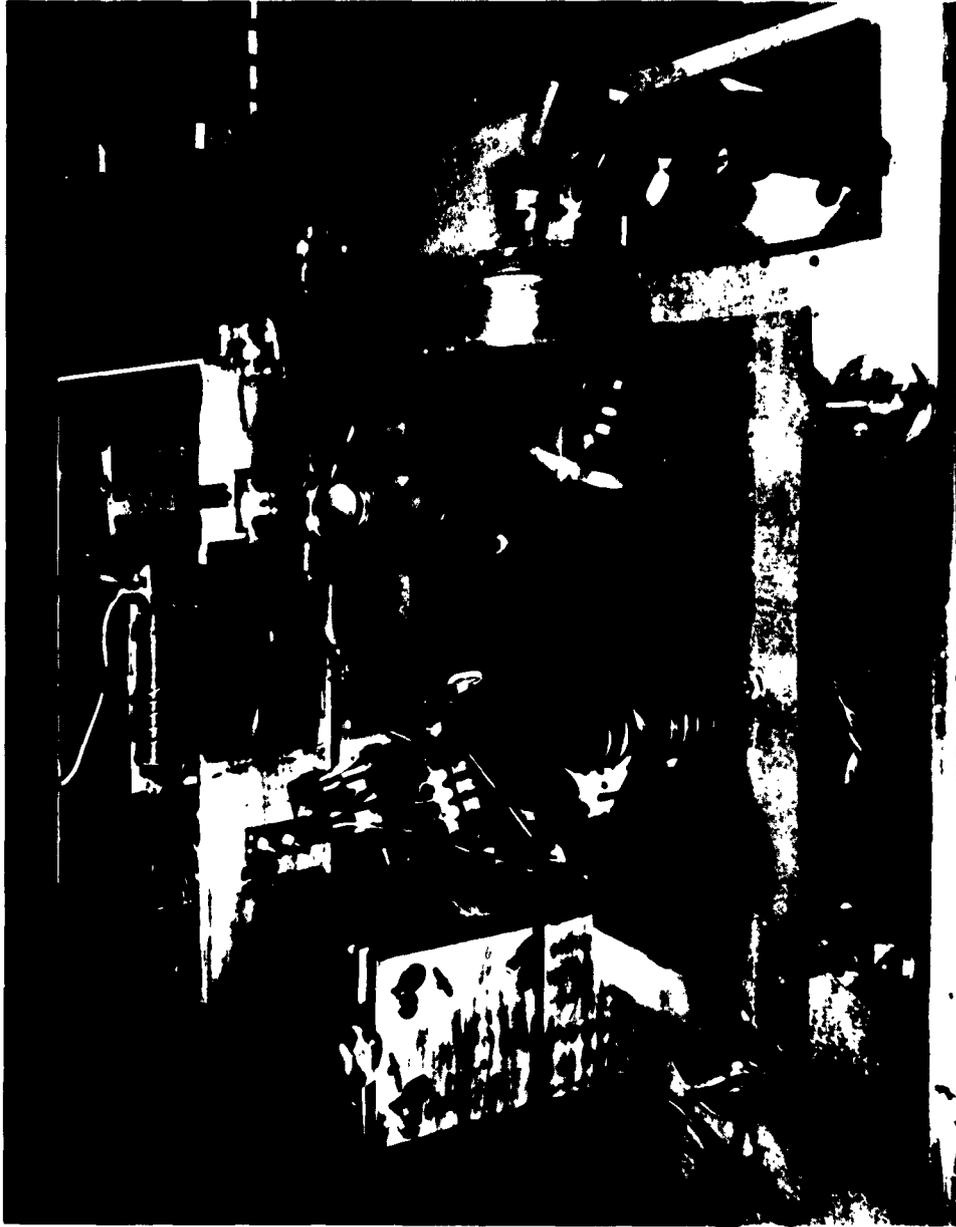
Fig. III-2. The far infrared spectrometer. The 1-meter gas cell is shown in the sample space. The specular reflection attachment has been removed (see Fig. III-1) and is shown at lower right in this picture.

reflection measurements at a 45° angle of incidence. Q_2 and Q_3 provide the use of two reststrahlen plates crossed at the polarizing angle in the manner suggested by Czerny[8] and Strong.[9] This device is described in a paper by Lord and McCubbin.[4]

Below 50 cm^{-1} no reststrahlen filters are available, and the filtering rests upon zero-order gratings.[10,11] These gratings are designed to diffract out of the beam radiation of just shorter wavelength than that required, and the longer wave radiation sees the filter gratings as a series of mirrors and is specularly reflected. Measurements are being made on filter gratings with the lines crossed in three dimensions. This arrangement appears to show considerable improvement in reducing the amount of stray radiation over an arrangement with two gratings in the same plane with the lines parallel to the main grating, G.

With wide slits, carefully chosen zero-order gratings, and the improved detector, the long-wavelength limit should be extended out beyond 600 microns with high hope of reasonably pure radiation.

For use with the low-temperature bolometer, the kinematically mounted mirror, K, allows the beam to traverse the reflection filters or sample holders, R_1, R_2, R_3, and R_4, (two pairs similar to Q_2 and Q_3). The beam is brought to a focus in the tube, M, outside the instrument, and a 25-cm gas cell can be filled or emptied without disturbing the spectrometer. Polyethylene windows form a seal for the cell at each end. The 90° off-axis ellipse, E, provides a 6:1 image reduction and permits location of the detector, B, and cryostat, H, out of the light beam.[12]

The speed of the instrument is approximately f/4 and the usable slit height with the bolometer is 18 mm (12 mm with the Golay detector). Spectra are obtained by automatic recording, the grating being rotated slowly at a rate determined by the resolution and signal-to-noise ratio required. A smoothing circuit is used in the final stage before the pen recorder to give a selection of time constants up to 40 seconds to increase the S/N ratio. With a 25-second time constant approximately 1.5 hours is required to scan the range of one grating — approximately one octave. An automatic system can be used to reverse the sense of rotation, and can be set to allow the instrument to cycle over any particular wavelength range.

In Fig. III-3 the spectrometer that is being used in conjunction with the liquid-helium cryostat is shown (center of picture).

3. Sources

A study of plasma discharges has been carried out to ascertain whether a better source than the conventional H-4 mercury arc could be found for the far infrared. Equipment in the laboratory of the Plasma Physics Group of the Research Laboratory of Electronics was utilized. The initial experiment was a study of a dc plasma discharge in a magnetic field, and an optical system was designed and constructed so that this source
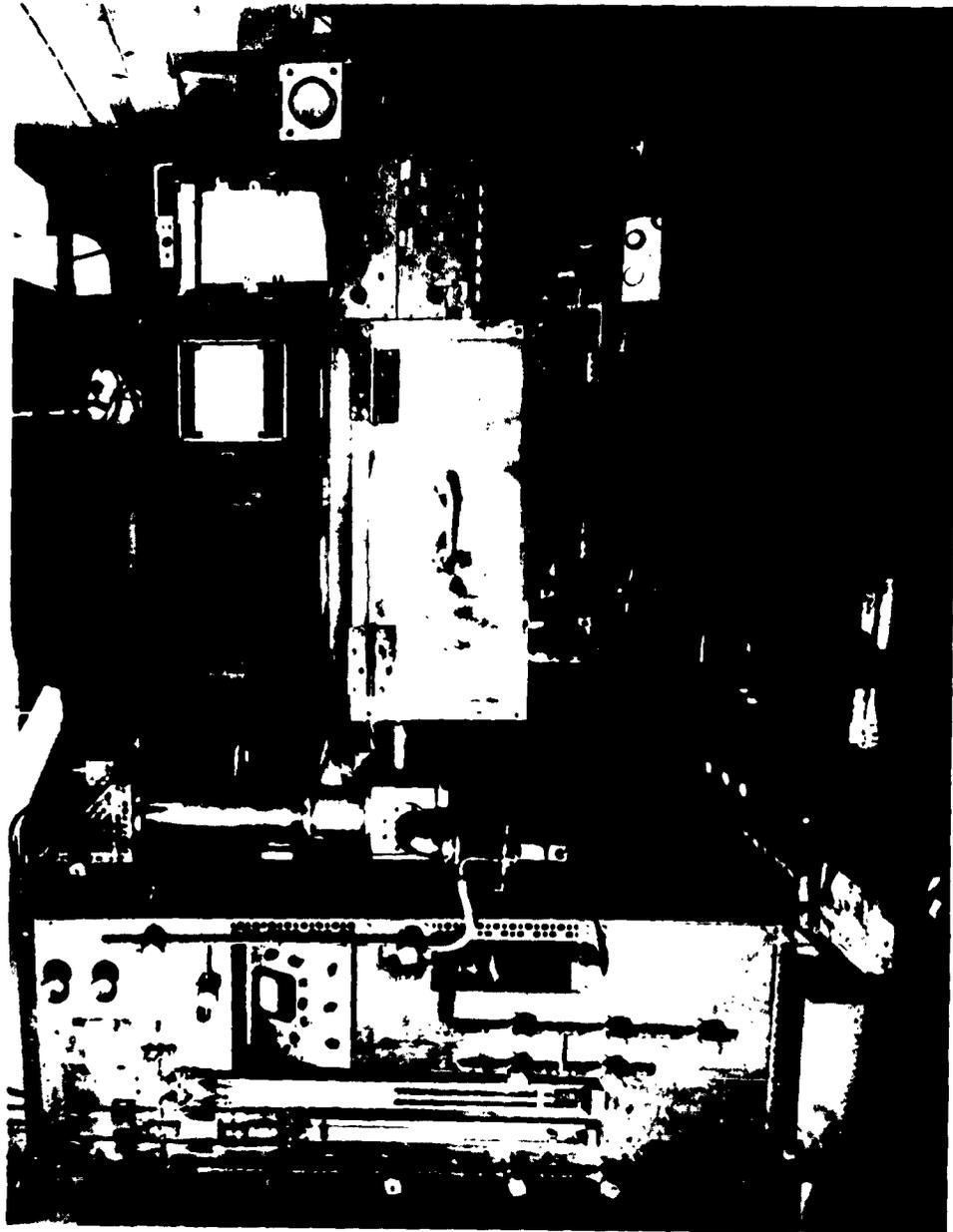
Fig. III-3. Arrangement of helium cryostat, spectrometer, and electronic equipment.

could illuminate a grating monochromator with radiation in the range 150-300 microns. A Golay cell (shielded from the magnetic field) served as detector, and two 6.4 ℓ/mm crossed replica gratings were employed as short-wave reflection filters. Additional transmission filters were used to eliminate radiation below 150 μ (black polyethylene, crystal quartz, mica, and a mixture of TlBr and KBr suspended in polyethylene). The source optics were arranged so that an f/2 beam was picked up from the source.

The design difficulties were increased by the fact that the plasma discharge lay along the axis of two Helmholtz coils, the space between which was only approximately 4 inches. The source consisted of a Pyrex tube of 3-in. diameter with a 1-in. crystal quartz window cemented onto a small side arm through which the radiation was collected. The argon discharge was 15 in. long and had a visible diameter of approximately 4 mm. The arc current was varied from 5 amps to 40 amps, and the focusing field from 0 to 1500 gauss. The pressure of argon was varied over the range 1-10 μ Hg, and the flow rate from 30 cc/min to 180 cc/min. None of these factors appeared to affect the far infrared signal within the experimental error of the detecting system. When the discharge was stopped, the signal did not immediately fall, and we concluded that most of the radiation was coming from the hot walls of the Pyrex tube and not from the plasma discharge.

In a second series of experiments a greater depth of the plasma discharge was examined in the absence of a focusing field. Here the discharge tube was 15 in. long with a quartz window on one end, from which the radiation was fed to the spectrometer. The argon-gas pressure was again varied and also the discharge current, but the far infrared signal appeared to be coming mainly from the hot walls of the tube.

A longer discharge, "1 meter" in length, containing four separate cathodes has been constructed and the Plasma Physics Group proposes to vary the discharge current up to approximately 40-50 amps, and to raise the pressure to 10-20 mm Hg in order to make the plasma more opaque. They also propose to repeat the first experiment, using fields up to 50,000 gauss with a simplified optical system in which only transmission filters will be used to eliminate short-wave radiation.

The results of the source studies, thus far, have been very disappointing.

4. Detectors

Most far infrared spectrometers use Golay detectors, which operate at room temperature. The limitation on the detectivity of these devices is set by the intrinsic noise of any device operating at 300°K. Improvement in detectivity beyond this limit can thus be sought only by going to lower temperatures.

During the past year we have placed in operation a liquid-helium cryostat in which an infrared-sensitive detector element can be refrigerated to 1.5°K. The cryostat was designed to be usable with any spectrometer of conventional design, and at present is

coupled with the far infrared vacuum spectrometer as shown in Figs. III-1 and III-3. Radiation from the spectrometer is fed to the detector at 1.5°K through a series of three windows at approximately 300°K, 77°K, and 4°K. (See Fig. III-4.) The window
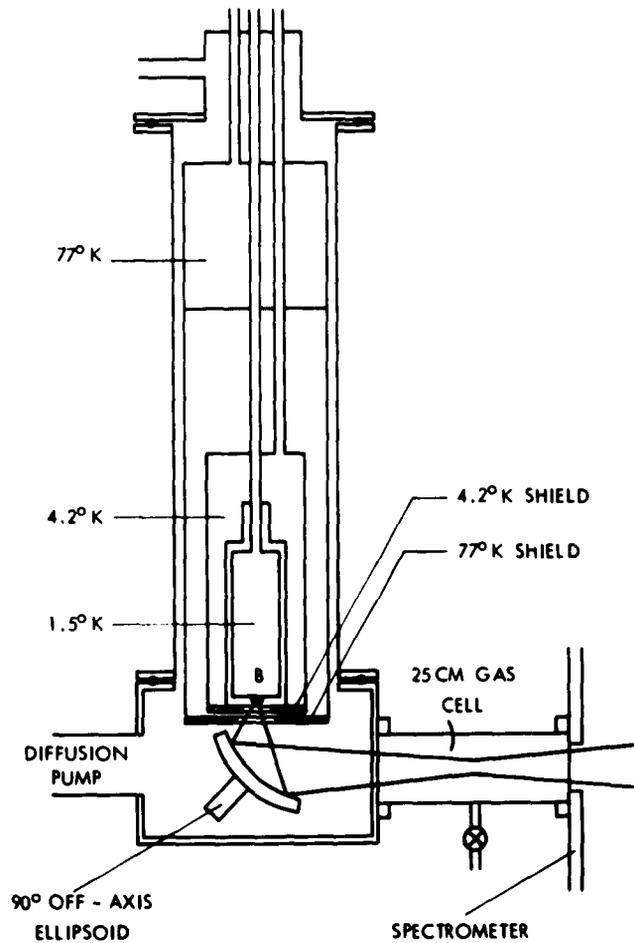
Fig. III-4. Detector optics and cryostat assembly.

materials are chosen to be opaque to radiation of wavelengths less than 40 microns. Thus the detector sees only a small part of the background radiant energy emitted by objects at room temperature in its field of view.

At present, the detector is a graphite bolometer of a fairly standard design.[13] It is operated at 1.5°K with radiation chopped at 13 cps. The spectrometer is arranged so that the radiation can be sent either to a standard Golay detector or to the cryostat. With this arrangement the performance of the low-temperature bolometer has been

compared quantitatively with that of the Golay detector.

Figures III-5 and III-6 show a comparison of the two detectors under the same operating parameters except for temperature. The gain of the Perkin-Elmer Model 107 amplifier was set to give the same throw of the pen on the recorder chart for the same slit widths, electronic time constants, and scanning speeds. Hence, by comparing the peak-to-peak noise, it is possible to assess the relative performance of the two detectors. The carbon bolometer was found to give a signal-to-noise ratio from 8 to 10 times better than the particular Golay detector employed. Boyle[13] gives a factor of 15 in a similar experiment, but the different figure is certainly within the spread in performance of different Golay detectors.

The improvement over the Golay detector was smaller than expected, but the low-temperature detector is considerably more rugged and stable than a Golay. It is possible to control the sink temperature within less than 0.01°K at 1.5°K quite conveniently by controlling the speed of pumping on the He, so that the zero-level drift is negligible. Continuous cycling between room temperature and helium temperature did not appear to affect the bolometers' performance, and they were quite stable over a particular scan. The detectivity remained approximately constant from run to run, provided the same bolometer current was used. This current was adjusted to give a maximum in the signal-to-noise ratio.

The current noise in the bolometers was comparable to or slightly greater than the amplifier noise, and there seemed to be no great advantage in cooling the load or the grid resistor of the preamplifier (Tektronix 122 or 123).

Recent measurements on a new type of semiconductor bolometer (single-crystal gallium-doped germanium) showed that it was advantageous to cool both input resistors to reduce Johnson noise. Preliminary measurements for far infrared detection on this type of bolometer proved quite promising. The properties of germanium are even better than carbon at low temperature, since the thermal conductivity of germanium is approximately three times that of carbon, and consequently the thermal relaxation time depends only on the thermal time constant of the leads. This means that higher chopping frequencies could lead to a reduction in 1/F noise and also to an increase in the scanning speed without loss of responsivity and consequent loss in resolution. Low[14] reported time constants as low as 0.5 msec with an N.E.P. of 5×10^{-13} watt at a chopping frequency of 200 cps.

The doped germanium absorbs about 70 per cent of the incident far infrared radiation. This result was obtained by making transmission and reflection measurements at the same time on a sample at a temperature less than 2°K with two carbon bolometers at right angles to one another and with the germanium specimen set at 45° to the incident beam. (See Fig. III-7.)

The germanium bolometer has an even lower response in volts/watt than the carbon
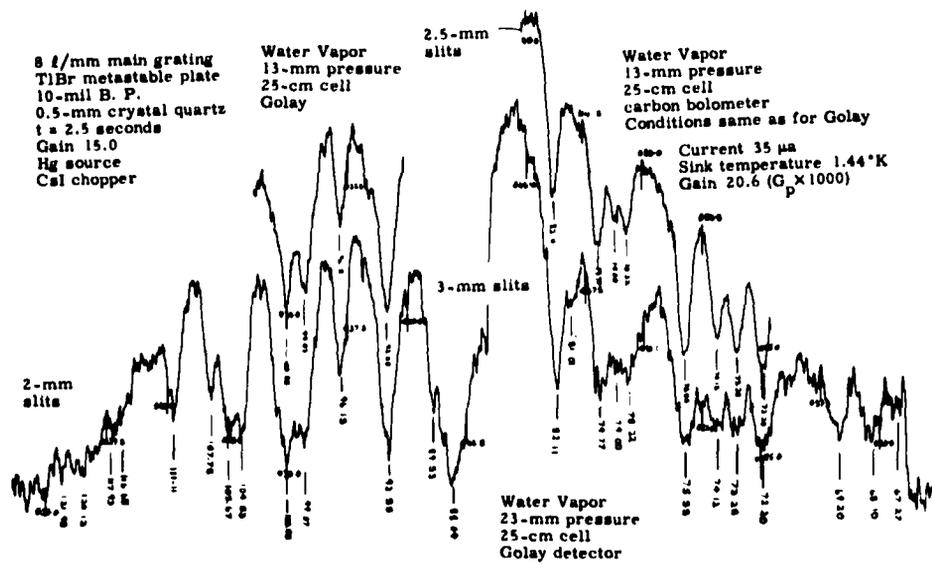
Fig. III-5. Infrared absorption spectrum of water vapor taken with carbon-resistance bolometer at 1.5°K (upper curve) compared with Golay detector at room temperature (lower curve). Note the clear resolution of the water triplet at 78.22, 79.00, and 79.77 cm^{-1} obtained with the low-temperature detector.
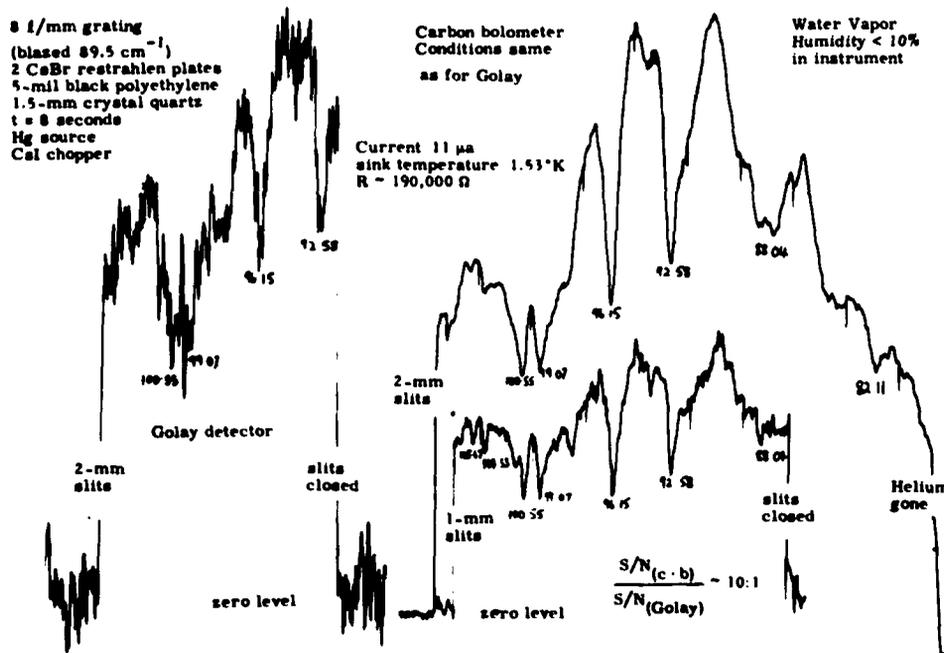
Fig. III-6. Comparison of signal-to-noise ratios for carbon bolometer (upper right curve) and for Golay detector (left curve). Lower right curve shows increased resolution with narrower slits for the carbon bolometer. Note small band now resolved on the side of the 100.55 cm^{-1} band as compared with the upper trace.
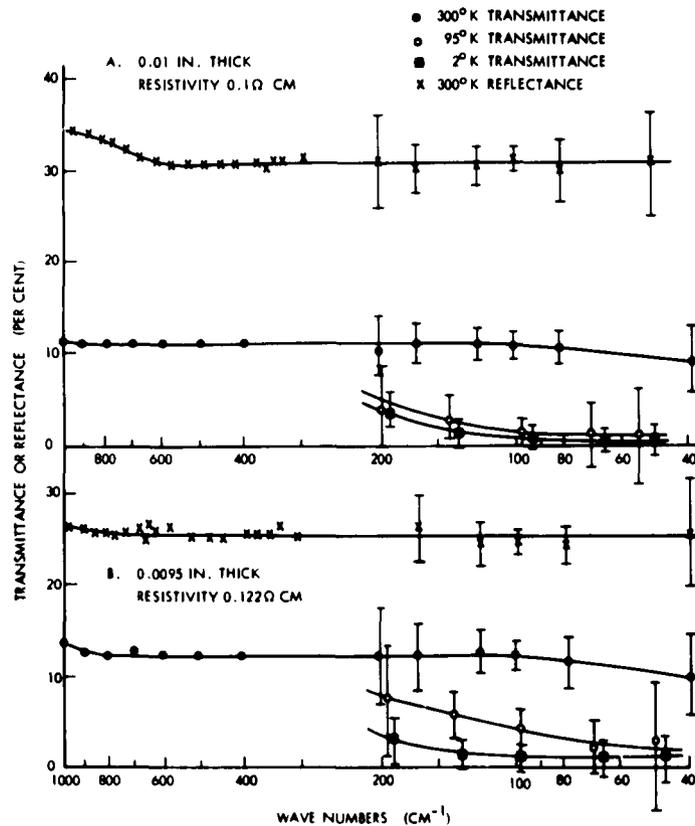
Fig. III-7.   Transmittance and reflectance curves for two gallium-
doped germanium bolometers $(1000\text{-}40 \text{ cm}^{-1})$.

bolometer and, although the inherent noise in the germanium bolometer is probably lower, more gain is needed in the preamplifier and the preamplifier noise (flicker noise in the tubes, etc.) may well set the lower limit of the minimum detectable signal.

Work will continue on the improvement of both types of bolometer. At present, their performance is less spectacular than we had anticipated, but, with continued effort, we hope to increase the factor of superiority over the Golay detector to the two orders of magnitude to be expected from the ratio of operating temperatures.

5. Filters

Considerable progress has been made during the year in the critical area of filter development. The most important filters have been the reflection filters.

For many years, the residual-ray reflection properties of crystals have been used to eliminate short-wave radiation. We have found for the first time that thallium
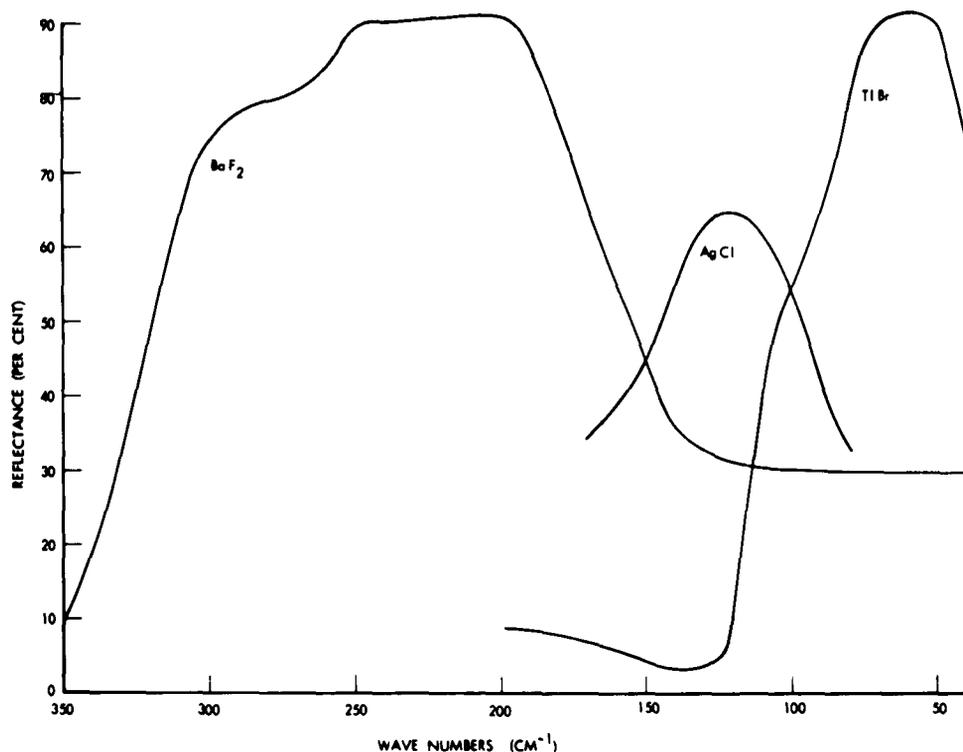
Fig. III-8. Reflectance curves for BaF_2, AgCl, and TlBr.

bromide removes radiation above 110 cm^{-1} quite effectively, yet has high reflectivity throughout the octave from 55 cm^{-1} to 110 cm^{-1}. Below 60 cm^{-1}, thallium bromo-iodide ("KRS-5") is somewhat better, although its lower limit is hardly less than that of TlBr. Silver chloride and barium fluoride have been found to reflect very well in the ranges 110-150 and 225-350 cm^{-1}, respectively. (See Fig. III-8.)

Below 50 cm^{-1} (down to approximately 20 cm^{-1}, the practical lower limit on our present equipment with the Golay detector), reflection filters made from a Bausch and Lomb magnesium grating of 4 grooves/mm and replicas of a 6.4-grooves/mm grating ruled at the University of Michigan have worked quite effectively. With such gratings used in the crossed orientation, less than 10 per cent stray radiation is present in the range 35-55 cm^{-1}. Comparable performance with these or other grating filters is expected below 35 cm^{-1}, but studies of stray radiation in this region are still under investigation.

B. SOME SOLID-STATE STUDIES IN THE FAR INFRARED

The infrared spectra of three perovskite titanates have been observed from 4000-30 cm^{-1} (2.5-330 μ) in both transmission and reflection. The spectra show a number

of absorption bands that can be interpreted as the frequencies of the spectroscopically allowed normal modes of vibration of these materials. In all three titanates an addtional infrared active vibration is observed which has not been reported previously. The value of the lowest vibrational frequency responsible for the ferro-electric behavior of these materials is estimated from a Kramers-Kronig (K-K) treatment of the reflection data. There appears, however, to be disagreement between reflection measurements made on these crystals at 5 cm^{-1} with millimeter wave techniques at Lincoln Laboratory, M. I. T., and our results down to 30 cm^{-1} with conventional infrared techniques. As the low-frequency band obtained from the K-K analysis is very dependent on the extrapolation to 0 cm^{-1}, the Lincoln Laboratory results will be repeated at 5 and 10 cm^{-1} before the data are complete.

Crystalline K_2PtCl_4, suspended in polyethylene, has also been studied in transmission from 400 cm^{-1} to 50 cm^{-1} (25-200 μ), and the allowed infrared active vibrations have been assigned. This work is being extended to other related compounds with the same crystal structure.

The author is grateful to Professor R. C. Lord for his advice and encouragement of this research, and he wishes to thank Dr. G. R. Hunt, Dr. B. N. Khanna, and other members of the Spectroscopy Laboratory, M. I. T., for their cooperation in this project.

C. H. Perry

References

1. T. K. McCubbin, Jr. and W. M. Sinton, J. Opt. Soc. Am. **40**, 537 (1950).

2. R. A. Oetjen, W. H. Haynie, W. M. Ward, R. L. Hansler, H. E. Schauwecher, and E. E. Bell, J. Opt. Soc. Am. **42**, 559 (1952).

3. L. Genzel and U. W. Eckhardt, Z. Physik **139**, 578 (1954).

4. R. C. Lord and T. K. McCubbin, Jr., J. Opt. Soc. Am. **47**, 689 (1957).

5. M. Yoshinaga, S. Fujita, S. Minami, A. Mitsuishi, R. A. Oetjen, and Y. Yamada, J. Opt. Soc. Am. **48**, 315 (1958).

6. T. J. Dean, G. O. Jones, D. H. Martin, P. A. Mawer, and C. H. Perry, Physica Suppl. **24**, 151 (1958).

7. D. Bloor, T. J. Dean, G. O. Jones, D. H. Martin, P. A. Mawer, and C. H. Perry, Proc. Roy. Soc. (London) **A260**, 510 (1961).

8. M. Czerny, Z. Physik **16**, 321 (1923).

9. J. Strong, J. Opt. Soc. Am. **29**, 520 (1939).

10. J. U. White, J. Opt. Soc. Am. **37**, 713 (1947).

11. R. F. Stamm and J. J. Whalen, J. Opt. Soc. Am. **36**, 2 (1946).

12. C. C. Helms, H. W. Jones, A. J. Russo, and E. H. Siegler, Jr., Spectrochimica Acta **19**, 819 (1963).

13. W. S. Boyle and K. F. Rodgers, J. Opt. Soc. Am. **49**, 66 (1959).

14. F. J. Low, J. Opt. Soc. Am. **51**, 1300 (1961).

IV. NUCLEAR MAGNETIC RESONANCE AND HYPERFINE STRUCTURE

Prof. F. Bitter
Prof. J. S. Waugh
Dr. L. C. Bradley III
Dr. H. H. Stroke
Dr. W. T. Walter
Dr. J. F. Waymouth
R. Arndt
G. S. Books

J. C. Chapman
A. T. Fiory
J. K. Flicker
T. Fohl
W. D. Halverson
E. R. Hegblom
D. J. Hegyi
R. G. Little
J. D. Macomber

H. C. Praddaude
O. Redi
C. J. Schuler, Jr.
R. W. Simon
W. W. Smith
M. J. Stavn
W. J. Tomlinson III
C. G. Wade

A. LEVEL CROSSINGS IN MERCURY 195* (40 hour) AND MERCURY 193* (11 hour)

New precision measurements of the hyperfine-structure interaction constants in radioactive mercury isotopes by optical detection of Zeeman level crossings[1] have been made. The magnetic fields at which the various crossings were observed are shown in Table IV-1, in terms of measured proton resonance frequencies. These values are still preliminary, since the analysis of systematic errors has not been completed. From the data in Table IV-1, the best values of the magnetic dipole interaction constants (A) and the electric quadrupole interaction constants (B) in the $6s6p^3P_1$ state are obtained (Table IV-2). The calculations were carried out on the IBM 7090 computer at the Computation Center, M.I.T., by using the program HYPERFINE 3-9.[2]

The ratio of the previously measured[1] interaction constants for Hg^{195} (9.5 hour) and Hg^{199} (both nuclear spin 1/2) has been calculated to be

$$\frac{A_{195}\left(^3P_1\right)}{A_{199}\left(^3P_1\right)} = \frac{f_p(195)}{f_p(199)} = 1.071925(63).$$

Combining this ratio with the recent direct measurement of the magnetic-moment ratio for these isotopes by Walter and Stavn,[3] we obtain a preliminary value for the Bohr-Weisskopf hyperfine-structure anomaly:

$$\Delta = 0.1466(85) \text{ per cent.}$$

This value for the anomaly will be refined by introducing single-electron interaction constants and second-order corrections to the energy levels, before comparing it with the value obtained from nuclear theory.[4]

The apparatus used in these experiments is shown schematically in Fig. IV-1, and the system assembly is shown in the photograph of Fig. IV-2. It is similar to that used by Hirsch.[5] Light from an electrodeless air-cooled Hg^{198} lamp (inside the scanning magnet) passes through a collimating lens and a quarter-wave plate-polarizer combination (which passes only one component of the Zeeman triplet when the light emerges parallel to the scanning field[6]) to a cell containing the vapor of the radioactive isotope
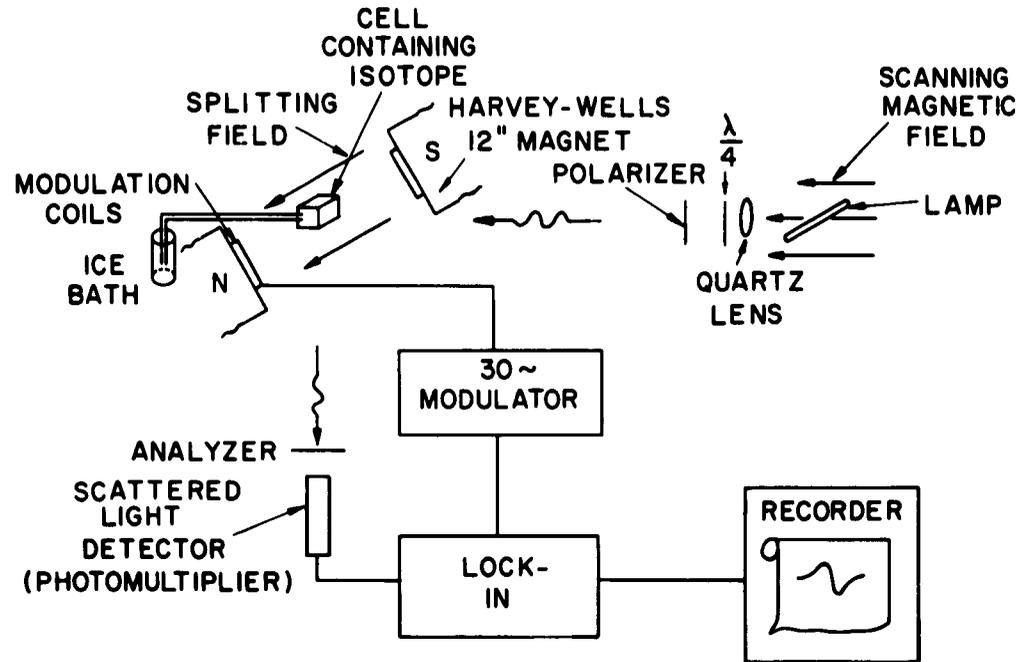
Fig. IV-1. Experimental arrangement for level-crossing work.

that is to be studied. The cell is in the gap of a Harvey-Wells 12-inch magnet of high homogeneity. The 2537 Å resonance radiation scattered from the vapor in the cell is monitored by a 1P28 photomultiplier and recorded as the "splitting field" of the Harvey-Wells magnet is slowly varied. There is a coherent contribution to the scattering from two Zeeman sublevels when they cross (become degenerate), at a particular value of the applied magnetic field, which changes the angular distribution of the scattered light. The resulting intensity resonance in the light scattered at 90° (a Lorentzian line of 1-10 gauss width) permits measurement of the crossing field to high accuracy (Table IV-1). Field modulation and phase-sensitive detection are used to increase sensitivity.

Field modulation was necessary in order to observe the second and third crossings in Hg^{195*}, since the intensity change at the crossings was less than 0.1 per cent. Although three attempts have been made, the corresponding crossings in Hg^{193*} have not been observed thus far, probably because of the presence of noise from other mercury isotopes in the cell.

In preparing for the Hg^{193*} experiment and in order to track down systematic errors in the Hg^{195} and Hg^{195*} measurements, a better way of measuring the magnetic field at a level crossing was worked out. At Professor Bitter's suggestion, the radioactive
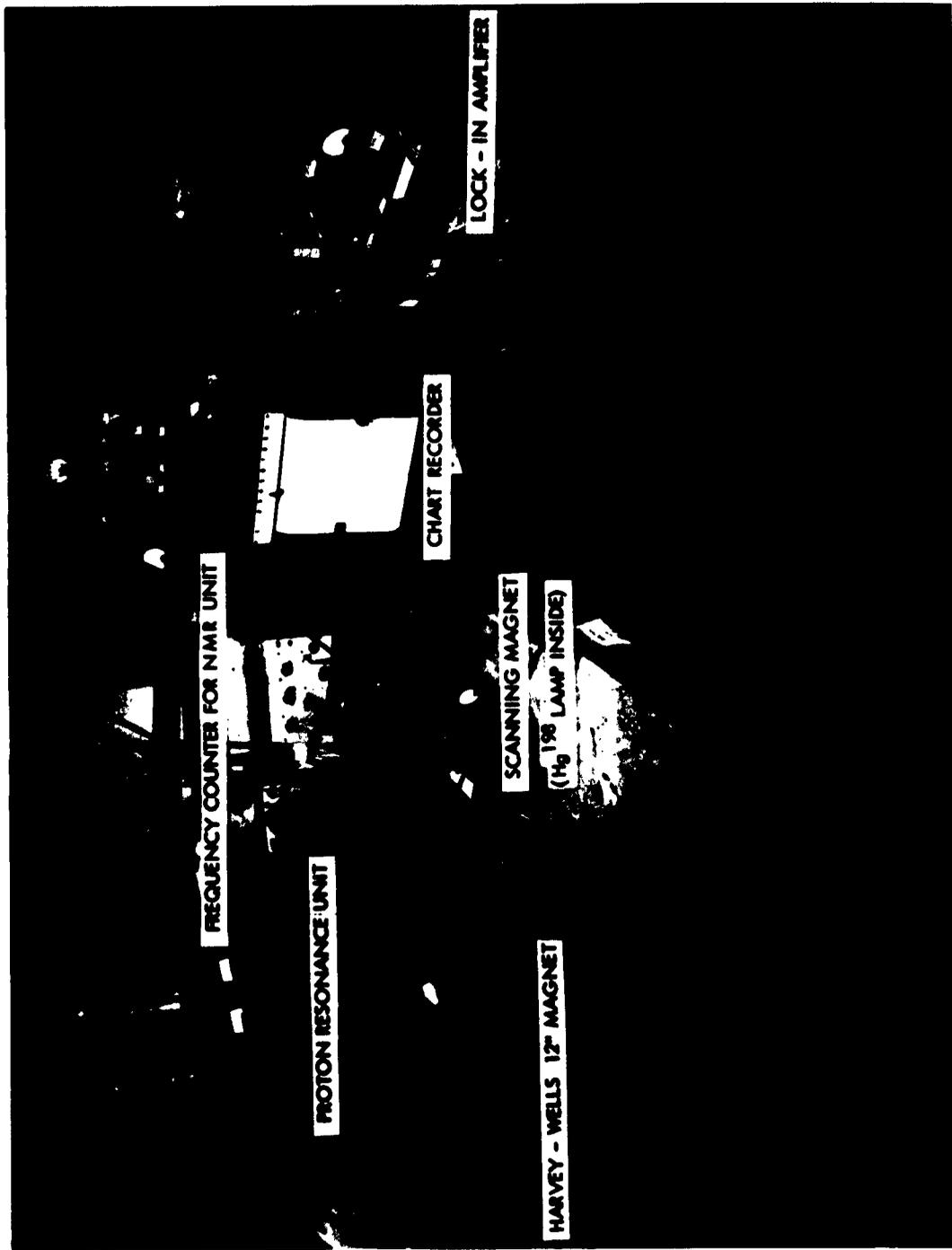
Fig. IV-2. The level-crossing apparatus.

Table IV-1. Proton resonance frequencies for level crossings. (Cell-to-probe correction, +1:25400, included in Hg^{195} and Hg^{195*} data.)

Isotope	Crossing	Frequency
Hg^{195}	$\begin{aligned} F &= \tfrac{3}{2}, \ m = -\tfrac{3}{2} \\ F &= \tfrac{1}{2}, \ m = +\tfrac{1}{2} \end{aligned}$	32371.3 ± 1.2 kc
Hg^{195*}	$\begin{aligned} F &= \tfrac{15}{2}, \ m = \tfrac{15}{2} \\ F &= \tfrac{13}{2}, \ m = \tfrac{11}{2} \end{aligned}$	33928.1 ± 1.9
	$\begin{aligned} F &= \tfrac{13}{2}, \ m = \tfrac{11}{2} \\ F &= \tfrac{13}{2}, \ m = \tfrac{7}{2} \end{aligned}$	32684.1 ± 2.7
	$\begin{aligned} F &= \tfrac{13}{2}, \ m = \tfrac{9}{2} \\ F &= \tfrac{13}{2}, \ m = \tfrac{5}{2} \end{aligned}$	31580.7 ± 21.0
Hg^{199}	$\begin{aligned} F &= \tfrac{3}{2}, \ m = -\tfrac{3}{2} \\ F &= \tfrac{1}{2}, \ m = +\tfrac{1}{2} \end{aligned}$	$30199.2 \pm .8$
Hg^{193*}	$\begin{aligned} F &= \tfrac{15}{2}, \ m = \tfrac{15}{2} \\ F &= \tfrac{13}{2}, \ m = \tfrac{11}{2} \end{aligned}$	34380.5 ± 2.0

Errors are 3σ

Table IV-2. Values of the hyperfine interaction constants.

Quantity	Spectroscopic Results[a]	Present Work
$A_{195}\left({}^3P_1\right)$	15838 ± 130 mc	$15813.4 \pm .4$ mc
$A_{195*}\left({}^3P_1\right)$	-2367 ± 7	$-2368.3 \pm .1$
$B_{195*}\left({}^3P_1\right)$	-794 ± 90	-788.0 ± 2.7
$\dfrac{A_{195}\left({}^3P_1\right)}{A_{199}\left({}^3P_1\right)} \approx \dfrac{f_p(195)}{f_p(199)}$		$1.071925 \pm .000063$
A_{193*}	-2398 ± 18	$-2400.1 \pm 1.0^{\mathrm{b}}$

[a]W. J. Tomlinson III and H. H. Stroke, see Section IV-B and Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, p. 18.

[b]Preliminary value: it is assumed that B is nearly the same as for Hg^{195*}.

cell was positioned slightly away from the center of the magnet gap, in the median plane between the cylindrical pole pieces. The proton resonance probe that measures the field was then placed in a symmetrical position on the opposite side of the gap center from the cell (see Fig. IV-3). The magnet homogeneity was sufficient so that this had little effect on linewidths, but the field difference between the cell and probe was considerably reduced. The residual correction (approximately 0.075 gauss) was less than the linewidth of the proton resonance signal. We found that it is possible to estimate this residual correction by observation of a beat pattern in the envelope of the magnetic-resonance side-wiggles; this pattern appeared when two proton resonance probes were connected in parallel and placed in the cell and in the probe positions. The probe at the cell position sees a slightly different field from the second probe. The beat pattern arises from the difference between the Larmor precession frequencies for the two samples. It is rendered visible on an oscilloscope with a calibrated time base. This beat effect, which I discovered independently, was, to my knowledge, first observed by R. Gabillard.[7]
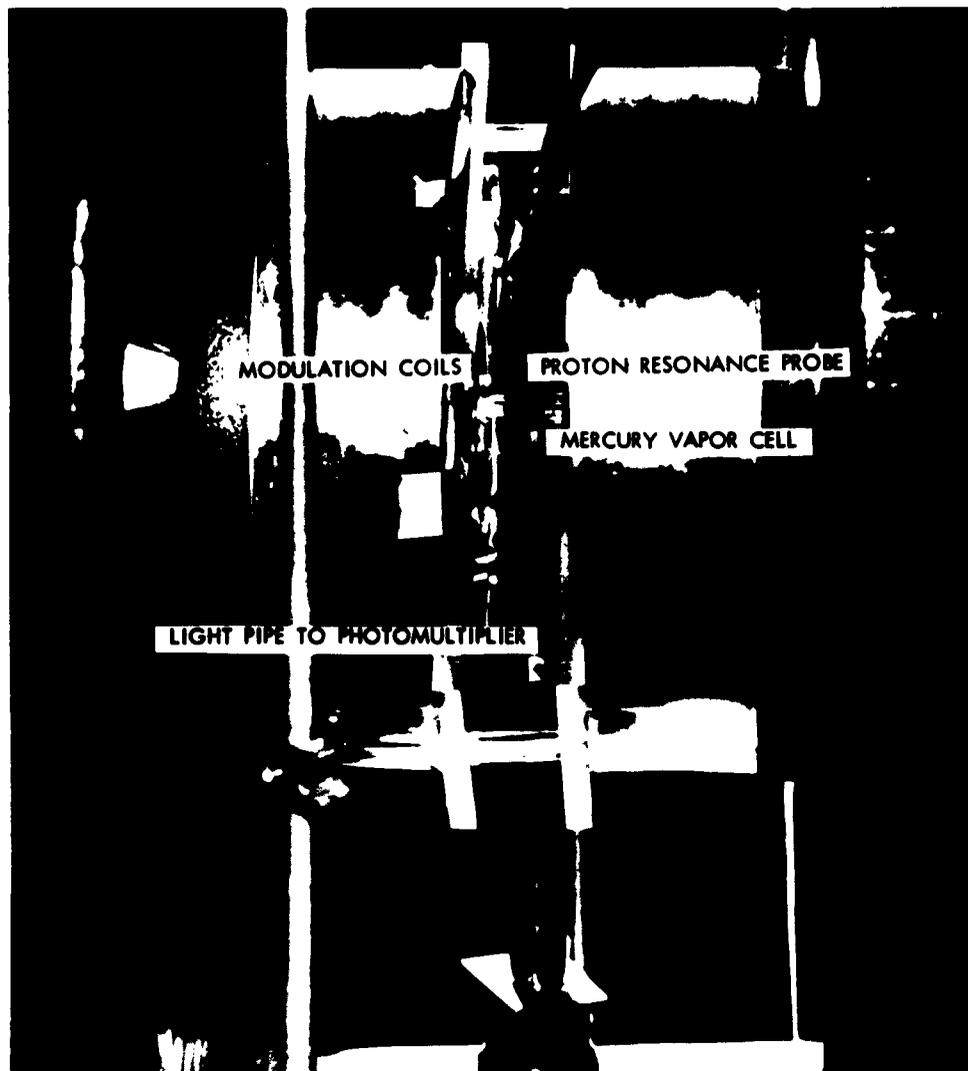
Fig. IV-3.   Close-up of magnet gap showing modulation coils, cell, and probe in position.

The position of the level crossing in Hg^{199} was remeasured with the cell and probe in symmetrical positions as a calibration. The results agree quite well with the very precise measurements of Kaul.[8]

A preliminary report of this work was presented at the American Physical Society 1963 Annual Meeting, in New York.[9] A final report will be submitted as a thesis to the Department of Physics, M.I.T., in partial fulfillment of the requirements for the degree of Doctor of Philosophy. We are grateful to A. Koehler and the Cyclotron Group, Harvard University, for the bombardments.

W. W. Smith

References

1. W. W. Smith, Level crossings in mercury 195 and mercury 195*, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M.I.T., October 15, 1962, pp. 28-29.

2. Obtained from Professor H. Shugart, Lawrence Radiation Laboratory, University of California, Berkeley, 1962.

3. M. J. Stavn, S.B. Thesis, Department of Physics, M.I.T., June 1963.

4. H. H. Stroke, R. J. Blin-Stoyle, and V. Jaccarino, Phys. Rev. 123, 1326-1348 (1961).

5. H. R. Hirsch, Ph.D. Thesis, Department of Physics, M.I.T., September 1960.

6. W. W. Smith, A compact magneto-optic scanning apparatus, Quarterly Progress Report No. 63, Research Laboratory of Electronics, M.I.T., October 15, 1961, pp. 83-85.

7. P. Girivet, La Résonance Paramagnétique Nucléaire (Centre National de la Recherche Scientifique, Paris, 1955), pp. 145-146.

8. R. D. Kaul, Ph.D. Thesis, Department of Physics, Case Institute of Technology, Cleveland, Ohio, 1963.

9. W. W. Smith, Bull. Am. Phys. Soc. 8, 9 (1963).

B. HYPERFINE STRUCTURE AND ISOTOPE SHIFT IN Hg^{193}, Hg^{193m}, AND Hg^{192}

The experimental work in our investigation of the hyperfine structure and isotope shift of the neutron-deficient mercury isotopes has been completed. Analysis of the results of the last few runs is still in progress, but we have obtained the following preliminary results:

Hg^{193} $\quad I = \frac{1}{2}$

$$A\left(^{3}S_{1}\right) \approx (800 \pm 30) \text{ mK}$$

Hg^{193m} $\quad I = \frac{13}{2}$

$$Hg^{193m} \quad A\left(^3P_1\right) = (-80.0 \pm .6) \text{ mK}$$

$$B\left(^3P_1\right) = (-23 \pm 15) \text{ mK}$$

$$A\left(^3S_1\right) = (-116.5 \pm 1.0) \text{ mK}$$

Isotope shifts relative to Hg^{196} (2537 Å line)

$$Hg^{193} = (180 \pm 30) \text{ mK}$$

$$Hg^{193m} = (240 \pm 15) \text{ mK}$$

$$Hg^{192} = (315 \pm 30) \text{ mK}$$

For Hg^{193m}, the A value for the $^3P_1$ state is in fairly good agreement with the results of Davis, Kleiman, and Aung,[1] and with those of W. W. Smith (see Section IV-A), but the B value is considerably different from that of Davis and his co-workers. The Hg^{193m} isotope shift is also in fairly good agreement with their values.[1] The data for Hg^{193} and Hg^{192} are not as accurate as those for Hg^{193m}, but the results tend to support our hypothesis concerning the causes of odd-even staggering in isotope shift.[2]

W. J. Tomlinson III, H. H. Stroke

References

1. S. P. Davis, H. Kleiman, and T. Aung, J. Opt. Soc. Am. 53, 506A (1963).
2. W. J. Tomlinson and H. H. Stroke, Phys. Rev. Letters 8, 436 (1962).

C. ERRATUM

In Quarterly Progress Report No. 68 (page 21), in the report entitled "Hyperfine Structure and Isotope Shifts in Neutron-Deficient Mercury Isotopes," the first sentence of the last paragraph should read:

"We have also obtained a corrected value for the Hg^{194}-Hg^{198} isotope shift in the 2537 Å line (0.280 ± .015) cm^{-1}."

W. J. Tomlinson III, H. H. Stroke

D. THE MAGNETIC MOMENT OF MERCURY 195 BY MEANS OF OPTICAL PUMPING

Optical pumping enables the techniques of nuclear magnetic resonance to be applied to dilute gases and vapors. The possibility of applying this technique to the

measurement of magnetic moments of radioactive nuclei was first demonstrated with Hg^{197} (65 hours) in our laboratory.[1] Orientation of a second radioactive mercury isotope, Hg^{195} (9.5 hours), by optical pumping has now been achieved.

A single quartz cell was filled with the stable isotope Hg^{199} ($I = 1/2$) and the 9.5 h radio isotope Hg^{195}. Nuclear magnetic resonance in the oriented vapor of each isotope was optically detected in the light scattered from the cell.

Resonances, 30-50 cps wide, were observed at 726 kc and 678 kc. The measured ratio $\mu_{195}/\mu_{199} = 1.070356(66)$. Using Cagnac's Hg^{199} result,[2] we obtain $\mu_{195} = 0.190813(12) \mu_H$, and $\mu_{195} = 0.532892(33)$ nuclear magnetons without diamagnetic correction. The errors quoted are three times the standard deviations. Complete details of this experiment are reported in a Bachelor's thesis submitted by Melvin J. Stavn to the Department of Physics, M.I.T., May 17, 1963.

W. T. Walter, M. J. Stavn

References

1. W. T. Walter, Ph. D. Thesis, Department of Physics, M.I.T., 1962; Quarterly Progress Report No. 65, Research Laboratory of Electronics, M.I.T., April 15, 1962, p. 39.

2. B. Cagnac, Ann. phys. 6, 467 (1961).

V. MICROWAVE ELECTRONICS

Prof. L. D. Smullin   Prof. A. Bers   R. P. Porter
Prof. H. A. Haus   Prof. L. J. Chu   H. M. Schneider
                              P. A. Mandics

A. HIGH-PERVEANCE HOLLOW ELECTRON-BEAM STUDY

Direct-current and radiofrequency interaction measurements have been completed
on both a cylindrical-cathode and a conical-cathode magnetron injection electron gun.
Detailed dc measurements on the cylindrical-cathode gun have been performed previ-
ously, and have been described by Poeltinger.[1] The guns described in this report were
operated with 1-μsec pulses at voltages up to $V_a = 8$ kv. All rf interaction measure-
ments were performed at f = 1119 mc. Figure V-1 is a schematic diagram of the beam
tester. The first cavity is provided with two coupling loops (not shown) for the
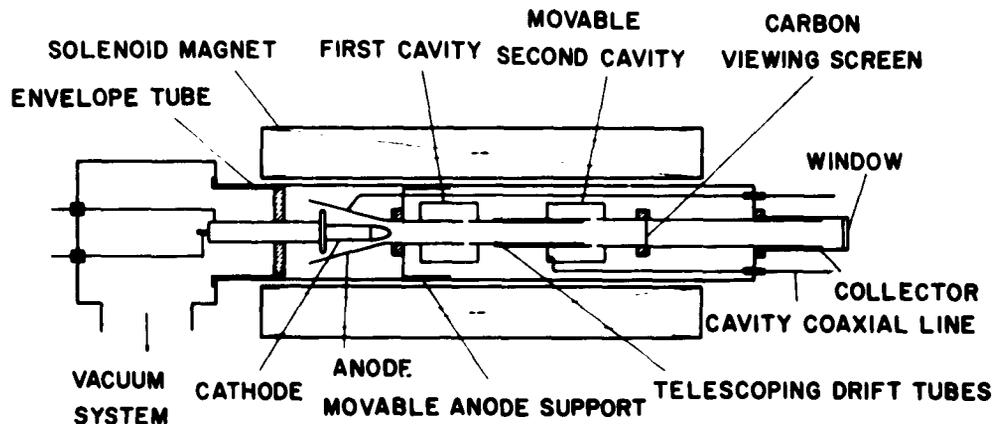
Fig. V-1. Schematic diagram of the beam tester.

measurement of beam loading. Figure V-2 gives the general layout of the cathodes,
cavity gaps, and mesh target for viewing the beam cross section.

1. Cylindrical-Cathode Gun

a. Measurement of the Beam Dimensions

The beam dimensions were determined by observing the heating of a car-
bonized nylon mesh screen that was placed in the path of the beam (see Fig.V-1).
Photographs of some typical beam cross sections are shown in Fig. V-3. These
are pertinent to the previously measured dc characteristics of a cylindrical-
cathode gun.[1]
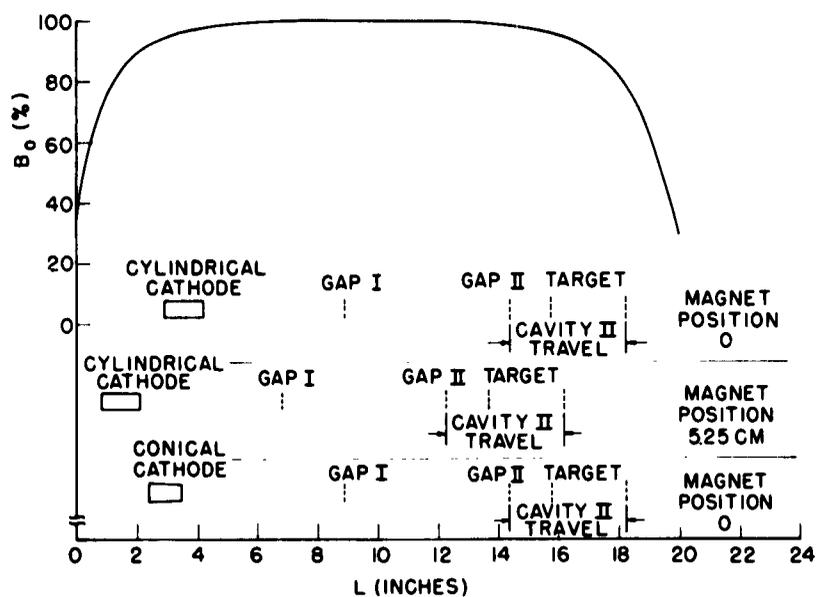
Fig. V-2. Relative magnetic flux density vs length of the magnet.
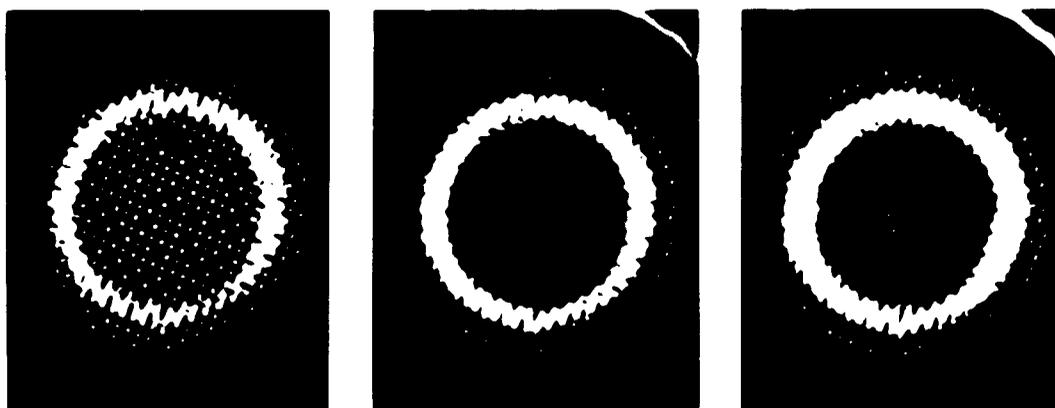
$U_o = 6$ KV	$U_o = 7$ KV	$U_o = 8$ KV
$I_o = 5.2$ A	$I_o = 6.4$ A	$I_o = 8.2$ A
$K_o = 11.2$ $\mu AV^{-3/2}$	$K_o = 10.9$ $\mu AV^{-3/2}$	$K_o = 11.5$ $\mu AV^{-3/2}$
$B = 1330$ G	$B = 1660$ G	$B = 1590$ G
$a = 0.98$ cm	$a = 0.98$ cm	$a = 0.98$ cm
$b = 0.81$ cm	$b = 0.81$ cm	$b = 0.84$ cm
$c = 0.66$ cm	$c = 0.65$ cm	$c = 0.62$ cm

Fig. V-3. Hollow-beam cross sections.

Results of beam cross section measurements for a new gun that was used to take all the rest of the data are summarized in the top of Fig. V-4. The accuracy of the
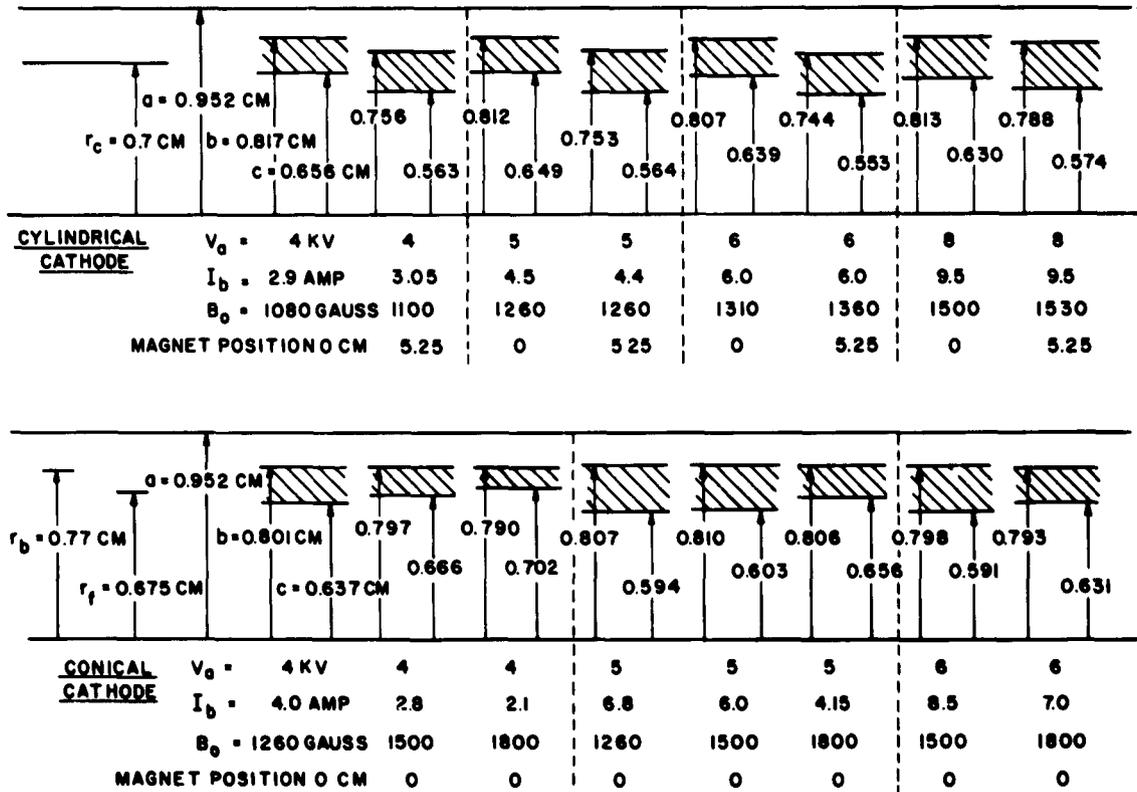
Fig. V-4. Beam dimensions.

dimensions obtained in this way was approximately 20 per cent.

b. Beam Loading

The real part of the electronic admittance G_{el} was measured by using the first cavity as a transmission cavity. The experimental arrangement for these measurements is diagrammed in Fig. V-5. G_{el} was calculated from the measured bandwidth of the cavity with and without the beam. These measurements, as well as the space-charge wavelength and gain measurements, were performed for two different magnet positions as shown in Fig. V-2. The measured and calculated dc parameters for all of the measurements are presented in Table V-1. The results of the beam-loading measurements are listed in Table V-2.

CAVITY WAVEMETER — 13-DB PAD

DETECTOR

240 Ω

PREAMPLIFIER — UHF GENERATOR — ISOLATOR — DIRECTIONAL COUPLER

OSCILLOSCOPE

PULSE GENERATOR AND AMPLIFIER

DETECTOR — VARIABLE ATTENUATOR — 20-DB PAD — TRANSMISSION CAVITY CAVITY I — HOLLOW BEAM

240 Ω
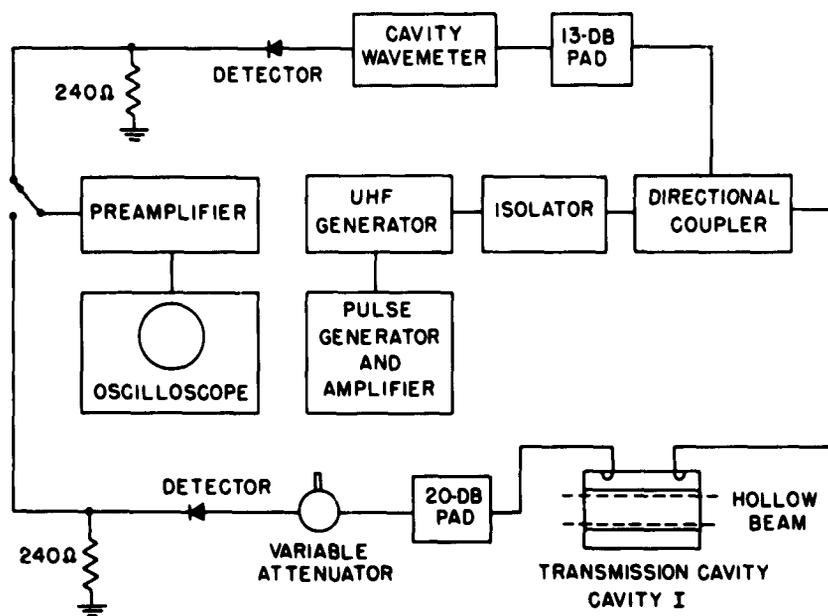
Fig. V-5. Beam-loading measurement.

Table V-1. D-C characteristics of the cylindrical-cathode gun.

Case	V_a (kv)	V_b (kv)	I_b (amp)	B_o (gauss)	Magnet Position (cm)	K (microperv)	f_p (mc)
1	2	1.92	0.8	1030	0	9.0	—
2	2	1.86	0.8	1030	5.25	9.0	—
3	4	3.80	2.9	1080	0	11.4	742
4	4	3.68	3.1	1100	5.25	12.0	743
5	5	4.77	4.5	1260	0	12.7	866
6	5	4.59	4.4	1260	5.25	12.5	854
7	6	5.70	6.0	1310	0	12.9	954
8	6	5.47	6.0	1360	5.25	12.9	960

Table V-2. Beam loading of the cylindrical-cathode gun.

Case	G_{el} Measurement ((ohm)$^{-1}$×10^6)	Calculation Kinematic Theory ((ohm)$^{-1}$×10^6)	Kinematic Theory Thin Beam (b≈c) ((ohm)$^{-1}$×10^6)	Space-Charge Theory Thin Beam (b≈c) ((ohm)$^{-1}$×10^6)
1	72.9	—	—	—
2	71.5	—	—	—
3	129	100	104	108
4	149	121	121	124
5	120	129	126	127
6	144	138	143	139
7	—	126	133	135
8	—	164	157	155
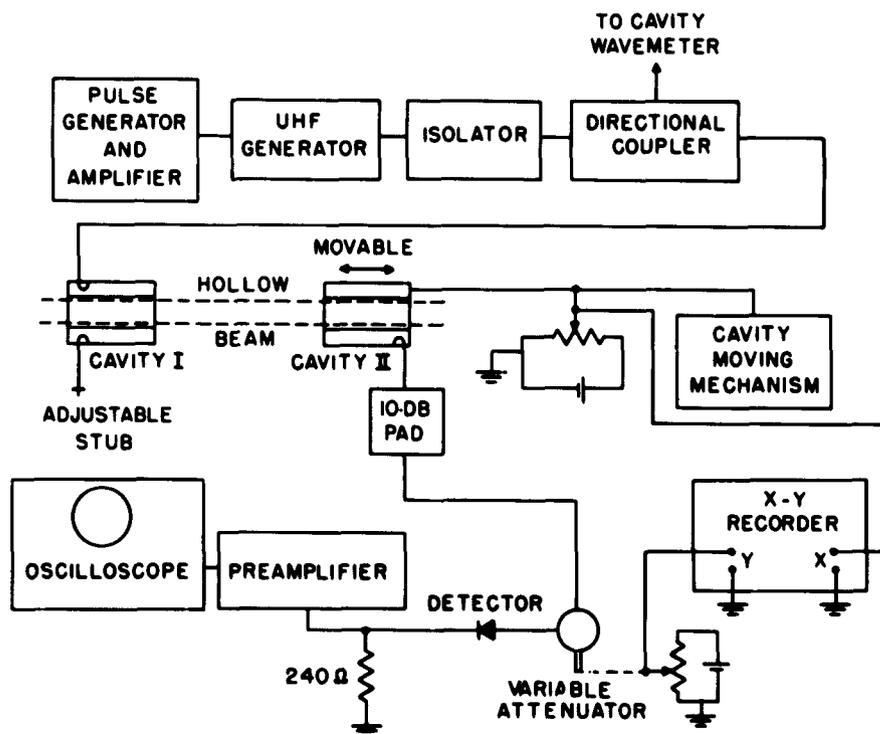
Fig. V-6. Space-charge wavelength, gain and noise measurement.

For purposes of comparison, G_{el} was calculated by using the approximate kinematic[2] and space-charge[3] theories based on a thin-beam assumption (b-c «c), and also by using the exact kinematic formulation.[2] The calculated values are also tabulated in Table V-2.

c. Space-Charge Wavelength and Gain

The space-charge wavelength and two-cavity gain were measured by using the arrangement illustrated in Fig. V-6. With a constant power input into the first cavity, the second cavity was moved along the beam and the power output from the second cavity was plotted against distance. The resultant curves are plotted in Fig. V-7.

The space-charge wavelength λ_q was calculated by making use of the plasma frequency-reduction factors of Branch and Mihran.[4] The available two-cavity gain was determined from the formulation given by Bers.[3] These theoretical values are compared with the experimental results in Table V-3.

d. Noise

The noise power output from the second cavity was plotted against distance along the beam for four differeit voltage settings. The circuit used for these measurements

Table V-3. Space-charge wavelength and gain of the
cylindrical-cathode gun.

Case	Measurement		Calculation			
	λ_q (cm)	Gain (db)	λ_q (cm)	Gain (db)	M^2	M^2Y_o $((ohm)^{-1} \times 10^3)$
1	17.1	7.8	17.5	—	—	—
2	15.0	6.2	13.8	—	—	—
3	18.2	9.4	19.7	10.5	0.545	1.16
4	11.9	2.6	14.8	4.74	0.458	0.701
5	18.9	10.0	19.5	13.1	0.593	1.45
6	13.3	3.9	15.6	7.4	0.511	0.913
7	19.2	8.5	20.6	13.4	0.627	1.58
8	11.8	4.0	16.4	6.1	0.554	0.92
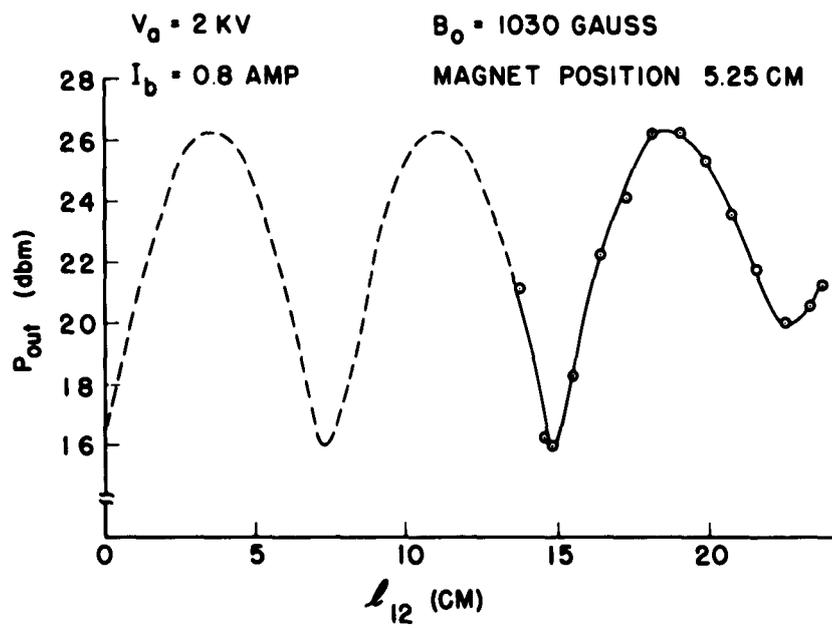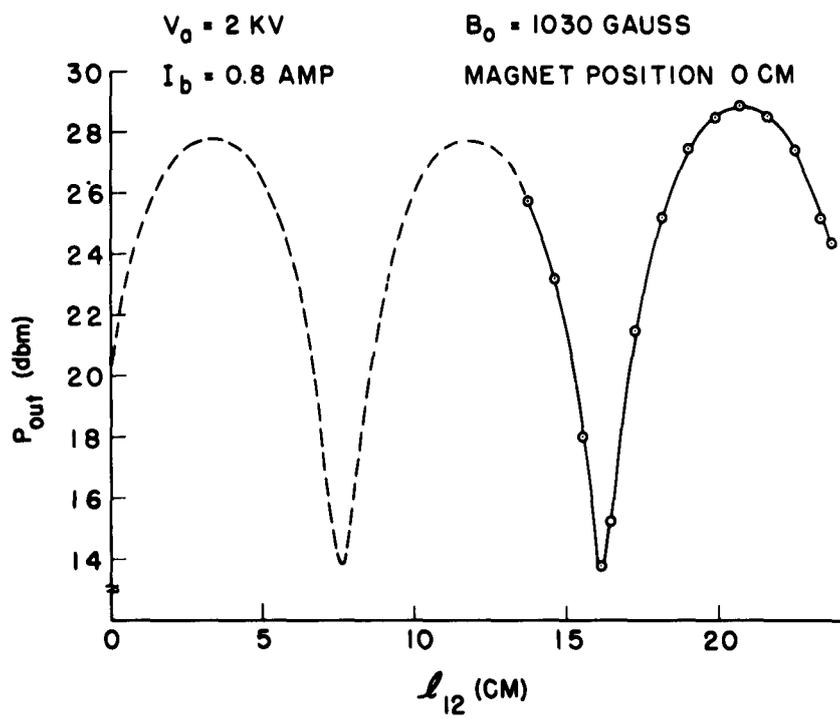
Fig. V-7. Second-cavity power output vs distance between cavity gaps for the cylindrical-cathode gun. (Cavity 1, P_{in} = 20 dbm.)
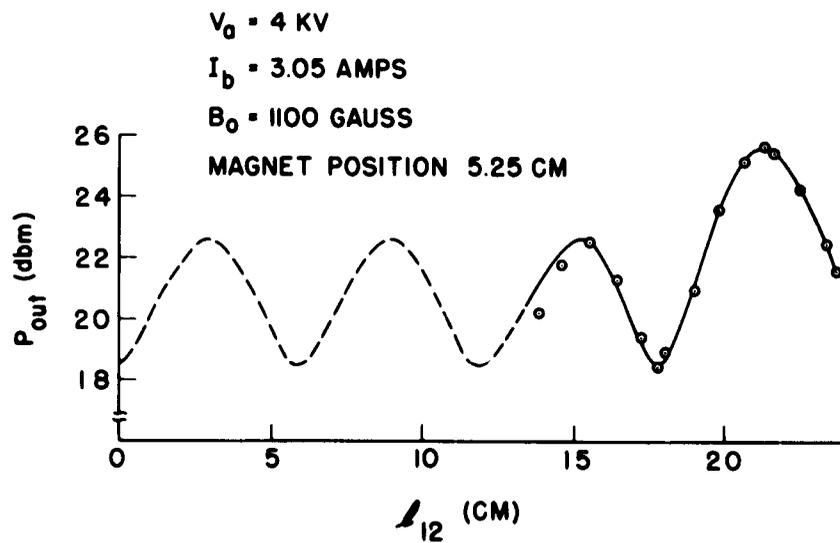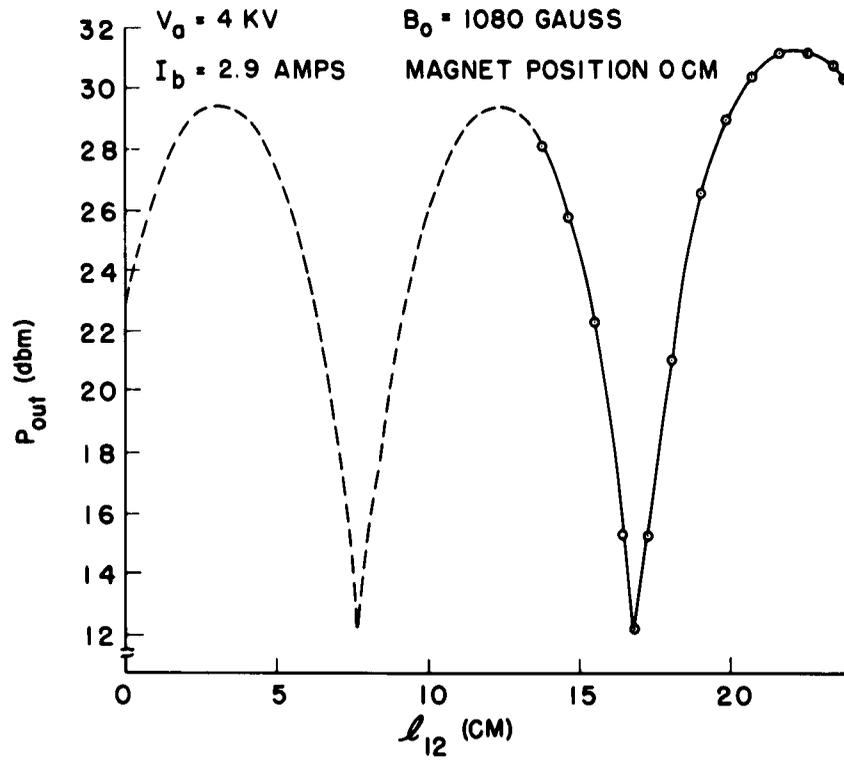
Fig. V-7. (continued).
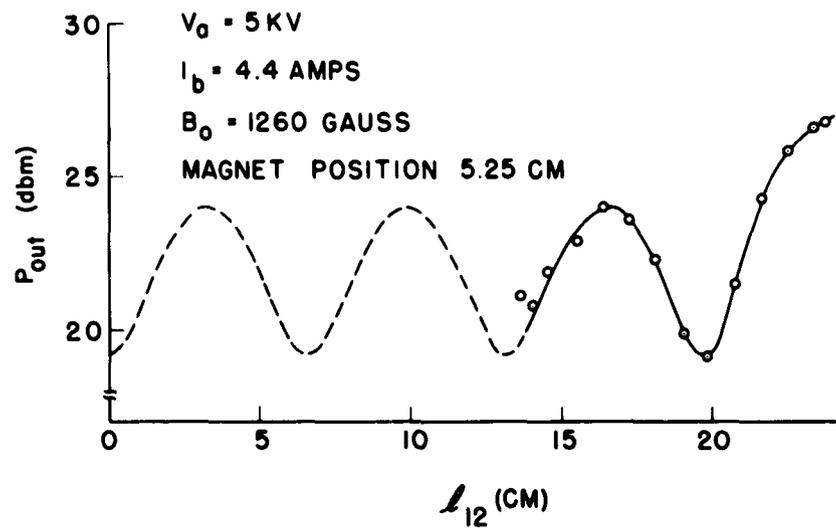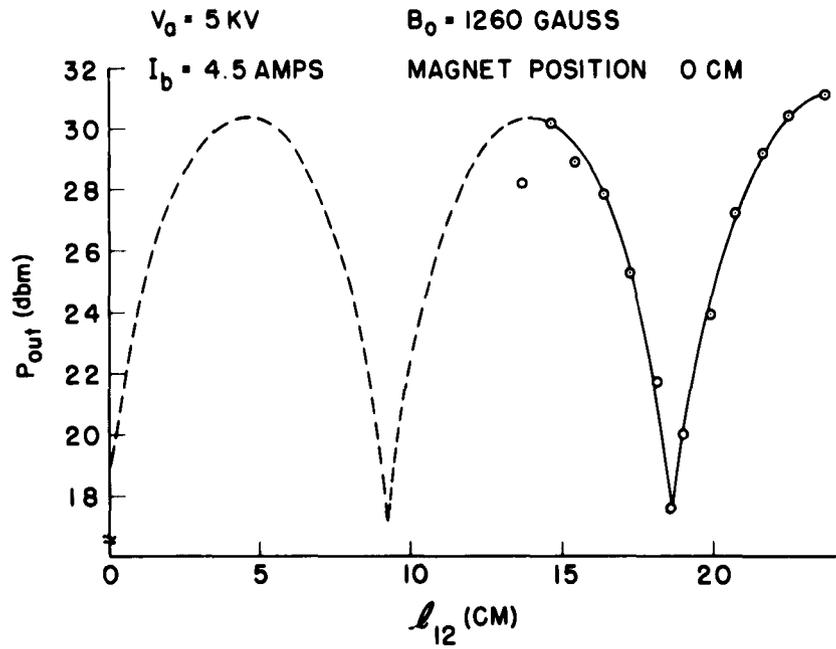
Fig. V-7. (continued).
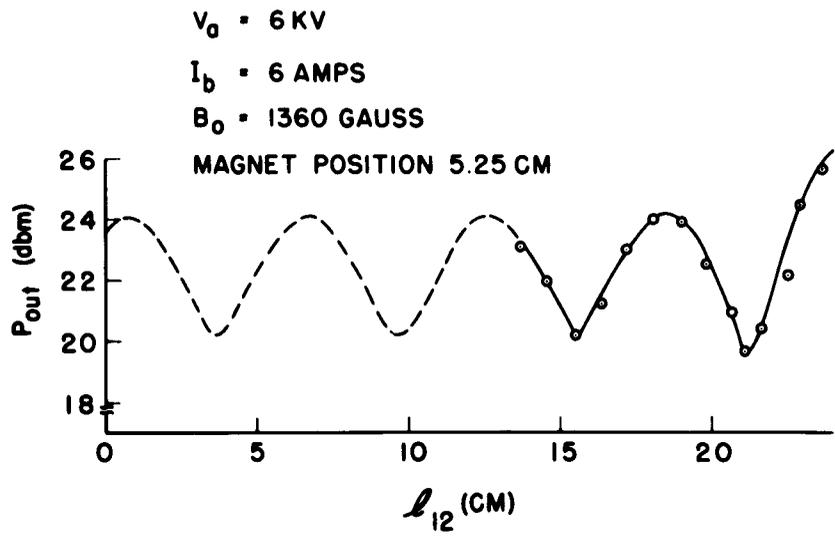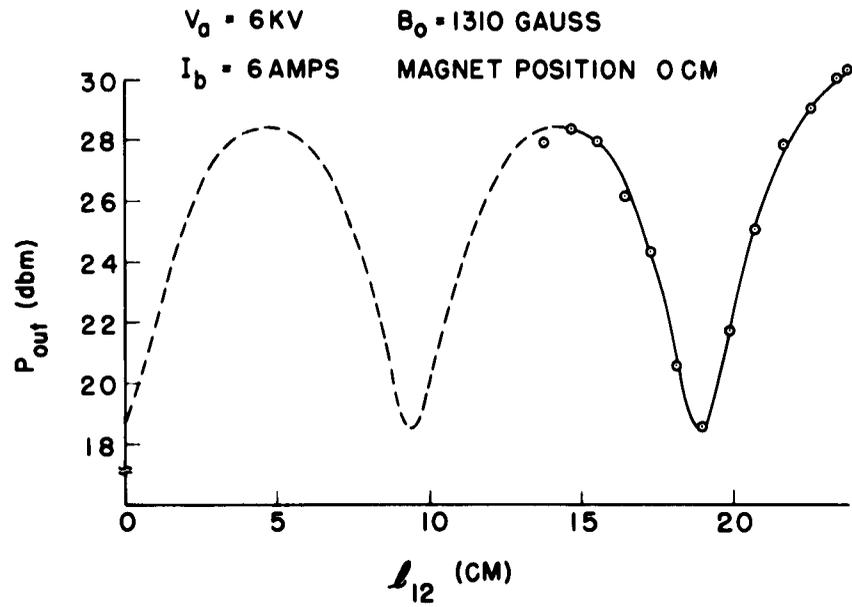
Fig. V-7. (concluded).
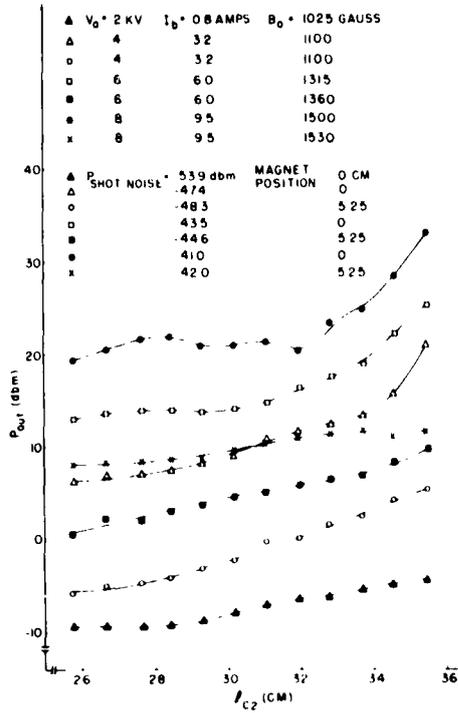
Figure V-8.

Second-cavity noise power output vs dis-
tance between cathode and cavity gap for
the cylindrical-cathode gun.

was identical to the one shown in Fig. V-6, with the exception that the power input to
the first cavity was disconnected. In Fig. V-8 the results of these measurements are
presented. It is quite apparent from Fig. V-8 that for a nonuniform magnetic field over
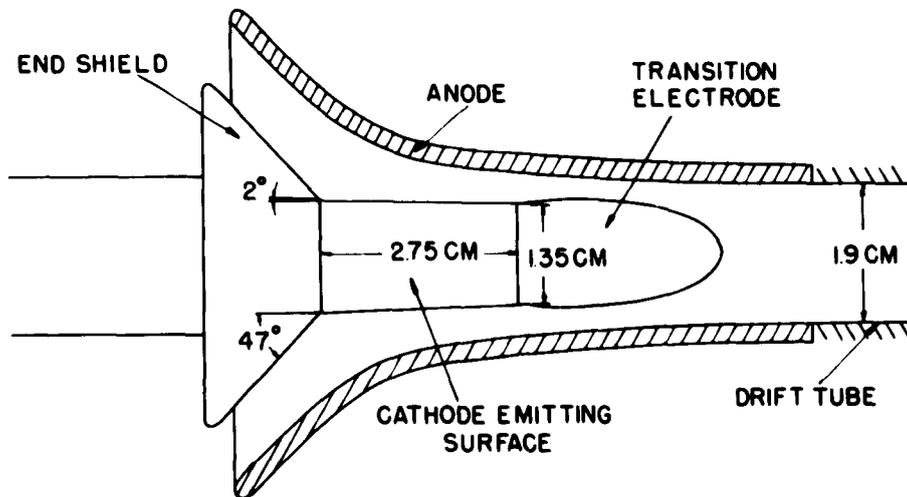
Fig. V-9. Conical-cathode magnetron injection gun.

the cathode corresponding to magnet position 5.25 cm (consult Fig. V-2) the noise output is drastically reduced. It was also observed that the noise output increased for an increase in the magnetic field. For all seven curves the noise output increases along the beam. In one case it increases by as much as 3 db/cm.

2. Conical-Cathode Gun

The design data and computer results for the electrode shapes of this gun have been described in a previous report.[5] After the cylindrical corrections[6] were carried out, the electrodes had the dimensions shown in Fig. V-9.

a. D-C Measurements

The results of the beam-dimension measurements are given at the bottom of Fig. V-4. In general, the beam cross section appeared to be quite symmetric. Beam breakup was observed only for low magnetic fields and high perveance, for example, for V_a = 7 kv, I_b = 10 amps, and magnetic field B_o = 925 gauss. The perveance $K = \dfrac{I_b}{V_a^{3/2}}$ is plotted against the magnetic field in Fig. V-10. For very high magnetic fields and relatively low voltages, V_a = 2, 3, and 4 kv, the perveance seemed to reach a limiting value of K = 7.5 microperv. On account of magnetic-field limitations, it was not possible to determine whether or not the higher voltage curves would also approach this (or some other) limit.
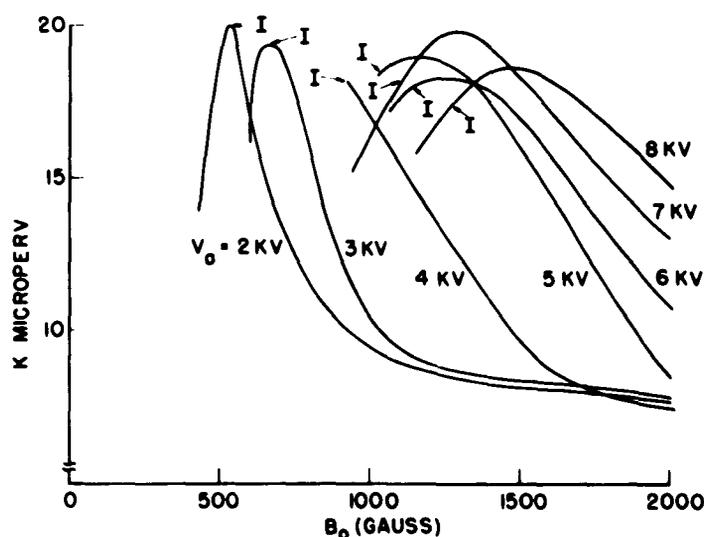
Fig. V-10. Beam perveance vs magnetic field for the conical-cathode gun.
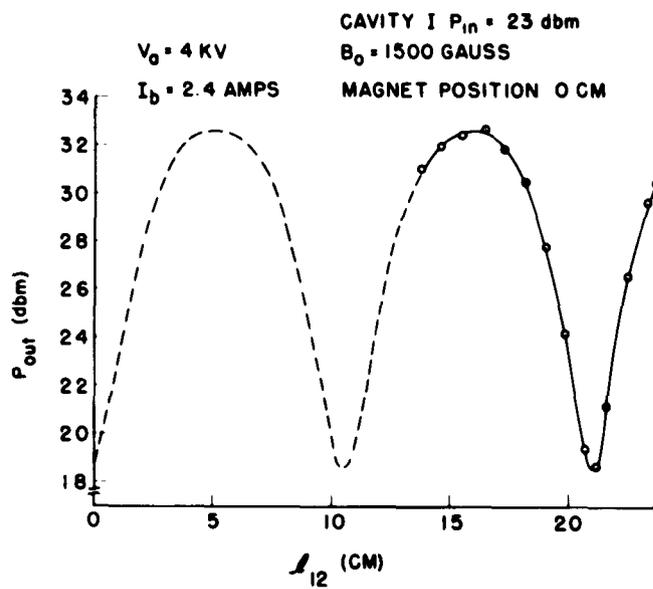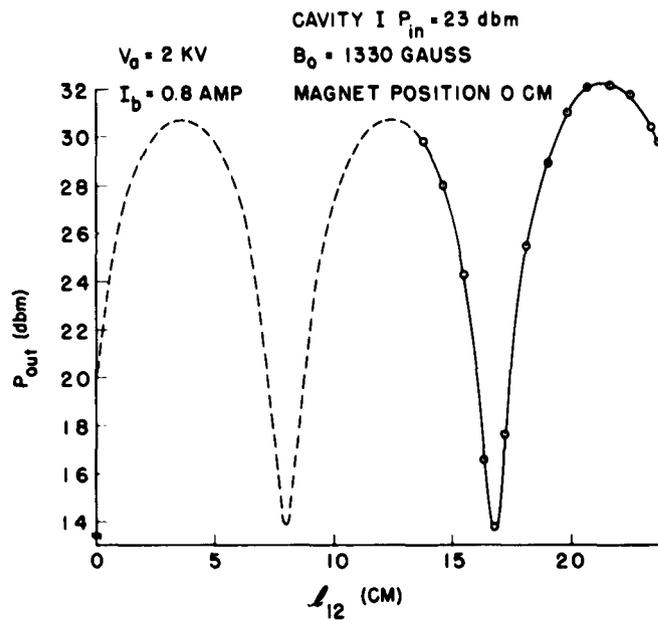
Fig. V-11. Second-cavity power output vs distance between cavity gaps for the conical-cathode gun.

Table V-4. D-C characteristics, space-charge wavelength, and gain of the conical-cathode gun.

Measurement

V_a (kv)	V_b (kv)	I_b (amp)	B_o (gauss)	Magnet Position (cm)	K (microperv)	λ_q (cm)	Gain (db)
4	3.79	2.4	1500	0	9.5	21.0	9.6

Calculation

λ_q (cm)	Gain (db)	M^2	$M^2 Y_o$ $((\text{ohm})^{-1} \times 10^3)$	f_p (mc)
21.8	13.1	0.552	1.13	744
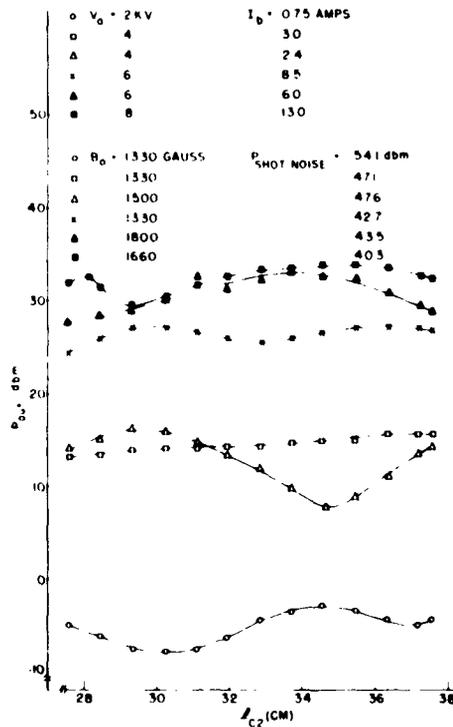
Fig. V-12. Second-cavity noise power output vs distance between cathode and cavity gap for the conical-cathode gun.

Note that the design parameters (V_a = 10 kv, K = 10 microperv, B_o = 1000 gauss) were not achieved. In fact, the gun gave much higher perveances than expected.

b. Cavity-Interaction Measurements

Space-charge wavelength and gain measurements were made for V_a = 2 and 4 kv as shown in Fig. V-11. For V_a = 4 kv, the measured and calculated values of λ_q and two-cavity gain are given in Table V-4. Because of the excessive amount of noise generated by the beam and the limited amount of power available at the input (P_{in} = 20 dbm) gain measurements could not be made for higher voltages.

The noise data presented in Fig. V-12 indicate no growth of noise power output along the beam. In fact, for the V_a = 2 and 4 kv cases, noise space-charge standing waves were observed with wavelengths equal to the ones measured with signal input to the first cavity. For both the cylindrical- and conical-cathode guns it was shown that the noise output from the second cavity (f = 1120 mc) had no relation to the noise observed on the collector current which was in the megacycle range.

P. A. Mandics, A. Bers

References

1. A. Poeltinger, High-perveance hollow electron-beam study, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, pp. 25-29.

2. A. Bers, Kinematic theory of gap interactions for relativistic electron beams, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, pp. 29-32.

3. A. Bers, Linear space-charge theory of gap interaction between an electron beam and electromagnetic fields, NTF 221, 53-60 (1961).

4. G. M. Branch and T. G. Mihran, Plasma frequency reduction factors in electron beams, Trans. IRE, Vol. ED-2, No. 2, pp. 3-11, April 1955.

5. A. Bers and A. Poeltinger, High-perveance electron-beam study, Quarterly Progress Report No. 64, Research Laboratory of Electronics, M.I.T., January 15, 1962, pp. 47-48.

6. A. Bers, L. Anderson, and K. Keller, Theory and Design of a Conical Electron Gun for Producing a Hollow Beam, Spencer Laboratory Engineering Report No. PT-277, Raytheon Company, Burlington, Massachusetts, 1962.

VI. MOLECULAR BEAMS[*]

Prof. J. R. Zacharias   R. S. Badessa   W. D. Johnston, Jr.
Prof. J. G. King   V. J. Bates   S. G. Kukolich
Prof. C. L. Searle   B. Brabson   F. J. O'Brien
Dr. G. W. Stroke   R. Golub   C. O. Thornburg, Jr.
                           G. L. Guttrich

A. CESIUM BEAM TUBE INVESTIGATION

During the past quarter, measurements of the frequency stability of the cesium atomic clock that was described in Quarterly Progress Report No. 69 (page 17) were made.
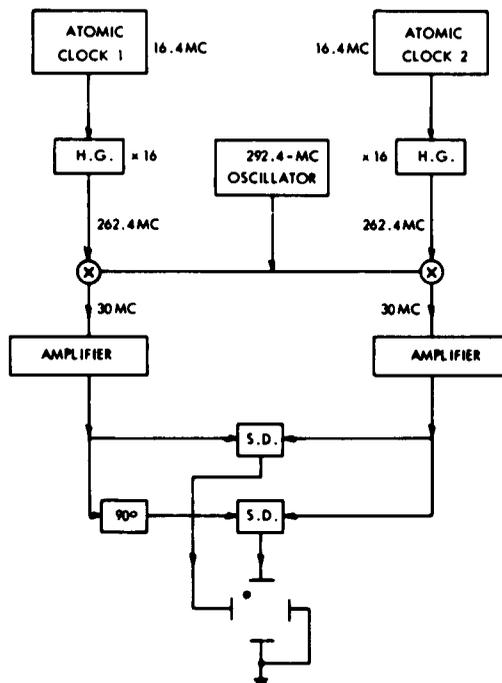
Fig. VI-1. Stability measurement system.

The system shown in Fig. VI-1 was used for these measurements. The 16.4-mc outputs of two identical clocks drive the harmonic generators which take the sixteenth harmonic. The resulting 262.4-mc signals are then mixed down to 30 mc by means of a common 292.4-mc local oscillator. Two synchronous detectors 90° out of phase provide a means of observing the instantaneous phase difference (or beat) between the two clocks. The

---

visual display consists of a dot on the oscilloscope which describes one complete circle for every sixteenth of a cycle of phase change at 16.4 mc. To check the stability, a record is kept of the variations in the time that it takes for some arbitrary number of complete cycles of the dot to occur. To permit measurements to be made unattended, an automatic data-recording system has been devised.

Measurements made on the system described above have shown that the stability of the clocks falls far short of the results that might be expected on the basis of the tests of the electronic apparatus described in Quarterly Progress Report No. 69 (pages 20 and 21). Since the instability must be caused by a component that would not influence the results of these electronics tests, it is probably caused by changes in the characteristics of the beam tubes themselves and, to some extent, by the modulators.[1] The variations in question consisted of a daily frequency fluctuation between the two clocks which was very closely correlated with the changes in the differential temperature between the two clocks during the course of a 24-hour period. A peak-to-peak variation of 3 parts in 10^{11} corresponded to a $10°F$ temperature change. Part of this variation was traced to the effect of temperature upon the 16.4-mc crystal oscillator and was eliminated by providing much more loop gain at very low frequency. However, an error of 2 parts in 10^{11} is still to be accounted for, and determination of the cause will be a major concern of the work during the next quarter.

R. S. Badessa, V. J. Bates, C. L. Searle

References

1. See Fig. III-2, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M. I. T., April 15, 1963, p. 18.

VII. RADIO ASTRONOMY[*]

Prof. A. H. Barrett   A. T. Anderson III   G. A. Garosi
Prof. J. W. Graham   J. C. Blinn III   J. W. Kuiper
Prof. R. P. Rafuse   R. K. Breon   B. R. Kusse
Dr. G. Fiocco   P. J. Brendel   D. H. Staelin
R. J. Allen   D. H. Steinbrecher

A. K-BAND RADIOMETRY AND OBSERVATIONS

The final results have been obtained for the observations of Venus, Taurus A, the Sun, and the Moon, made in December 1962. These are presented in Table VII-1.

Table VII-1. Results of observations at 1.18 cm wavelength, December 1962.

	Average Brightness Temperatures
Venus	$395 \pm {}^{75}_{55}$ °K
Sun	8870 ± 980 °K
Moon 3.5 days before full moon	240 ± 40 °K
0.3 days before full moon	254 ± 30 °K
1.8 days after full moon	254 ± 30 °K
	Flux density
Taurus A	275 ± 85 $\times 10^{-26}$ wm^{-2} cps^{-1}
	Attenuation
Zenith attenuation of the atmosphere	0.2 ± 0.1 db

The radiometer and experimental procedure were described in Quarterly Progress Reports No. 68 (pp. 35-36) and No. 69 (pp. 23-25).
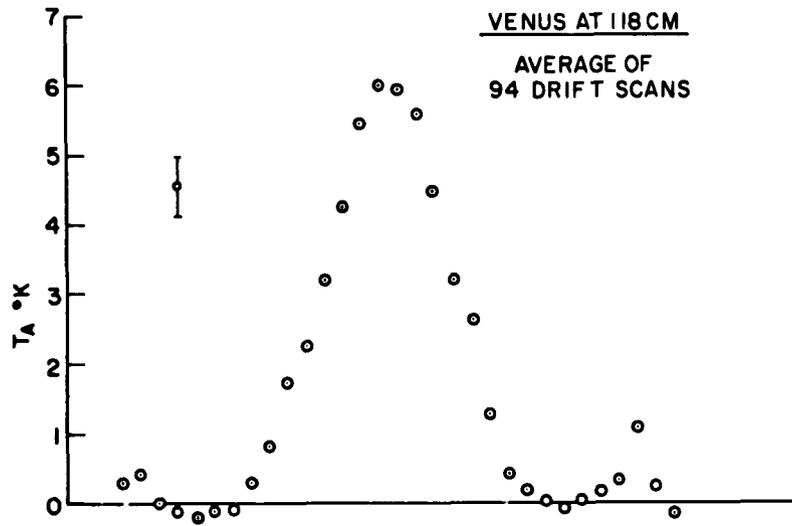
Fig. VII-1. Average of 94 drift scans of Venus.
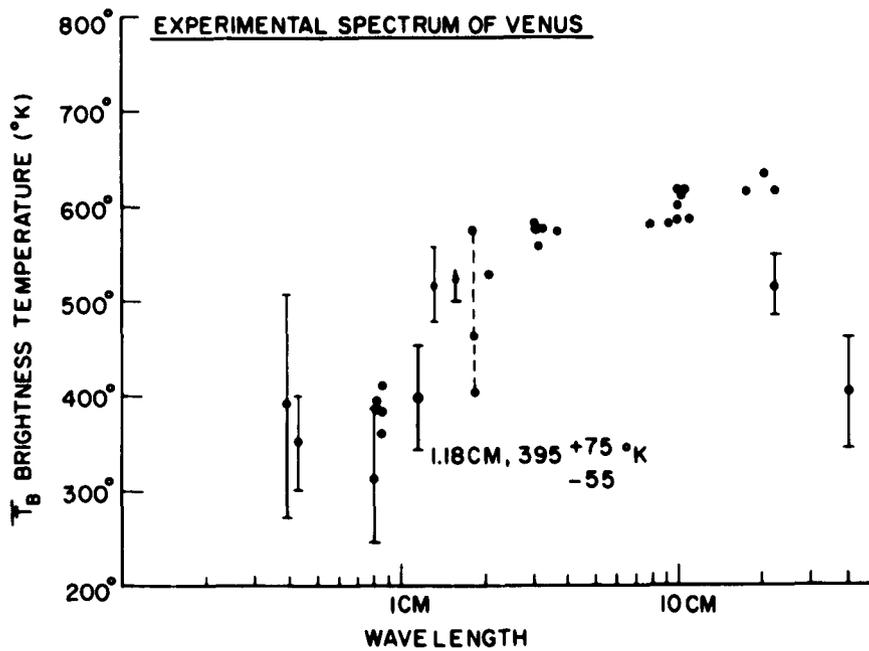
Fig. VII-2. Radiofrequency spectrum of Venus.

These values were computed by using an antenna gain derived from the gain measured at 35.2 kmc. This was only 0.2 db higher than the gain determined by using the solar and lunar data. The atmospheric attenuation was determined by brightness measurements of the sun and the sky as a function of elevation angle.

Figure VII-1 shows the average of 94 drift scans of Venus normalized to inferior conjunction. Figure VII-2 shows the final result for Venus and the results of other observers plotted as a function of wavelength. The sharp transition between Gibson and Corbett's measurement at 22.2 kmc and our measurement at 25.5 kmc may indicate the presence of resonant molecular absorption. A number of molecules have resonances in this frequency band. Measurements at several frequencies in this region are planned for the inferior conjunction in 1964, and any such strong features should be readily detected.

A. H. Barrett, B. R. Kusse, D. H. Staelin

B. A FOUR-MILLIMETER WIDEBAND MIXER

The noise figure of the 4-mm radiometer described in Quarterly Progress Report No. 69 (page 25), was 31 db, approximately 10 db greater than the design objective. The source of this difficulty was the crystal mixer. Development of an improved crystal mixer for use with this system is now complete, and the radiometer design objective of a 21-db noise figure has been realized.

Problems in the original mixer were rooted in the use of commercially available cartridge-type crystals designed for use with intermediate frequencies of a few tens of megacycles. Three odd picofarads of package capacitance shunting the I-F terminals of the crystal precluded suitable matching to the 50-ohm traveling-wave amplifier (TWT) I-F input over a 2-4 Gc passband.

It became apparent early that the answer to the I-F matching problem was to eliminate the diode package by building the diode in the waveguide. The diode whisker became the center conductor of a 151-ohm transmission line that was matched to 50 ohms with two quarter-wave transformers at the center of the I-F passband.

Since the I-F impedance of most mixer crystals is in the range 300-600 ohms, the 151-ohm output was too low to provide a 1:1 match, but did keep the missmatch below 2:1 over most of the I-F passband. VSWR measurements looking into the I-F terminals under operating conditions are shown in Fig. VII-3. The I-F impedance is strongly dependent on dc bias and local oscillator drive level, so that is was possible to get a very good match at the I-F port. However, the bias and drive levels for best I-F match did not agree with the levels giving the best system noise figure. Curves for each case are shown in Fig. VII-3.

Earlier work on harmonic mixers with microwave I-F outputs by Dr. J. C. Wiltse
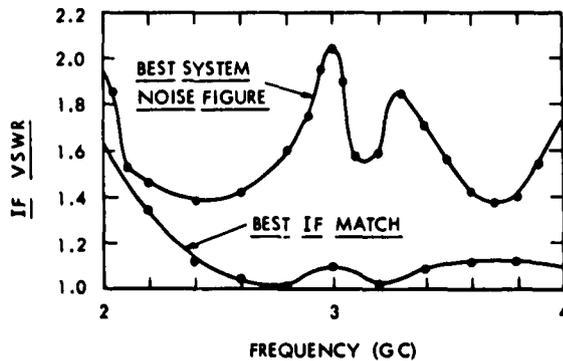
Fig. VII-3. VSWR at I-F port versus frequency.

and his group at Electronic Communications, Inc., Tinonium, Maryland, indicated that, of the semiconductors available, gallium arsenide was best suited for this application. Results of their experiments were to the effect that the I-F impedance of GaAs crystals could generally be made lower than silicon on germanium crystals without the necessity of excessively high drive levels. Since GaAs also has better high-frequency characteristics than silicon or germanium, it was used without additional investigation.

A completed mixer and its component parts are shown in Fig. VII-4. The whisker
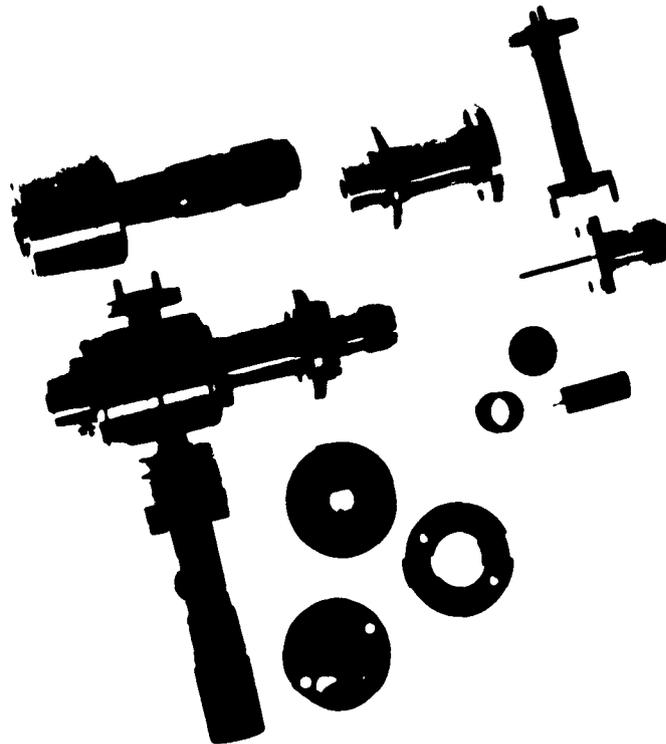
Fig. VII-4. Assembled mixer and component parts.
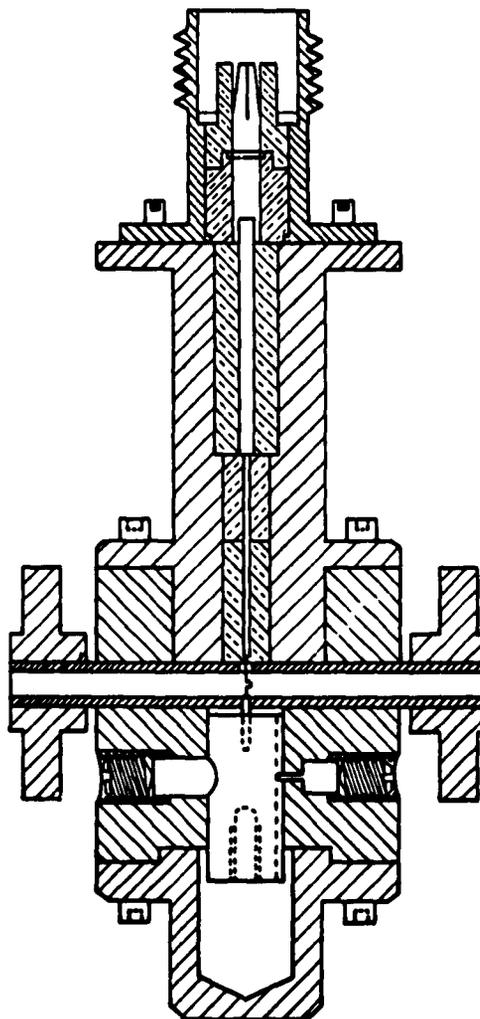
Fig. VII-5. Mixer assembly.

material was 0.001-inch phosphor bronze wire with an electroformed point. Radius of the contact area was of the order of 50 microinches. The whisker was an integral part of the center conductor of the matching section and passed through a 0.0135-inch hole in the top of a section of RG-99-U waveguide. The GaAs crystal was soldered on the end of a 0.020-inch silver post that passed through a 0.0205-inch hole in the bottom of the waveguide. A micrometer adjustment was used to position the crystal. Location of parts in the final assembly is illustrated in the cutaway view of Fig. VII-5. Forming was accomplished by repeated current pulses in the forward direction.

Design goals for mixer conversion loss, L_x, and noise-temperature ratio, t_x, were
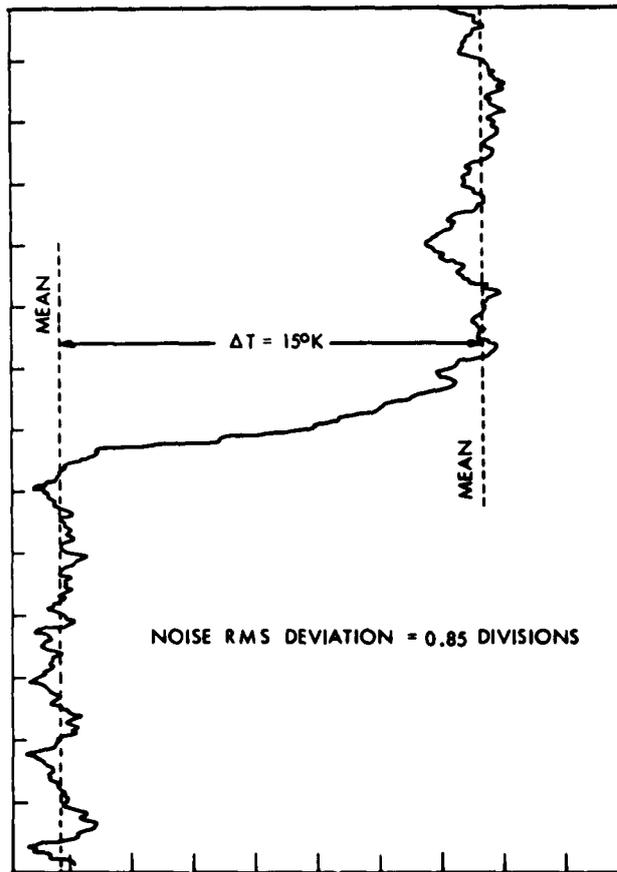
Fig. VII-6. Radiometer output with 4-sec integration time.

12 db and 2:1, respectively. For the mixers tested, L_x averaged approximately 12 db with the minimum around 9 db and the maximum around 13.5 db. Almost without exception t_x was less than 1.5:1.

The measured system noise figure for the mixer and the TWT S-band I-F amplifier, with a 5.5-db noise figure, was 15-16 db. The radiometer described in Quarterly Progress Report No. 69 (page 25), for which the mixer was designed, employs a narrow-band ferrite rf switch that prevents double sideband conversion. Measured radiometer noise figures were 19-20 db. Of the additional 4 db, 3 db were attributed to narrow switch bandwidth and 1 db to insertion loss between the mixer and the radiometer input.

The measured minimum detectable temperature for the radiometer was 1.5°K with an equivalent 1-sec integration time. This was well within the system design objective. A typical output record corresponding to an input temperature change of 15°K with an equivalent 4-sec integration time is shown in Fig. VII-6. The predetection bandwidth computed from measured minimum detectable temperature, radiometer noise figure,

and integration time was 1.8 Gc. This approaches closely the 2-Gc predetection band-
width that was the design objective.

D. H. Steinbrecher, J. W. Graham

C. A BROADBAND FERRITE SWITCH

A ferrite switch with a bandwidth equal to that of the waveguide has been designed.
It is expected to have an insertion loss of less than 1 db over the band 18-26 kmc, a
VSWR less than 1.5, isolation greater than 20 db, and a switching speed of approxi-
mately 10 μsec.

When the switch is on, the magnetic field is zero, and the wave travels unimpeded
through the switch with the wave polarization perpendicular to the resistive sheet (see
Fig. VII-7). When the switch is off, the axial magnetic field twists the polarization
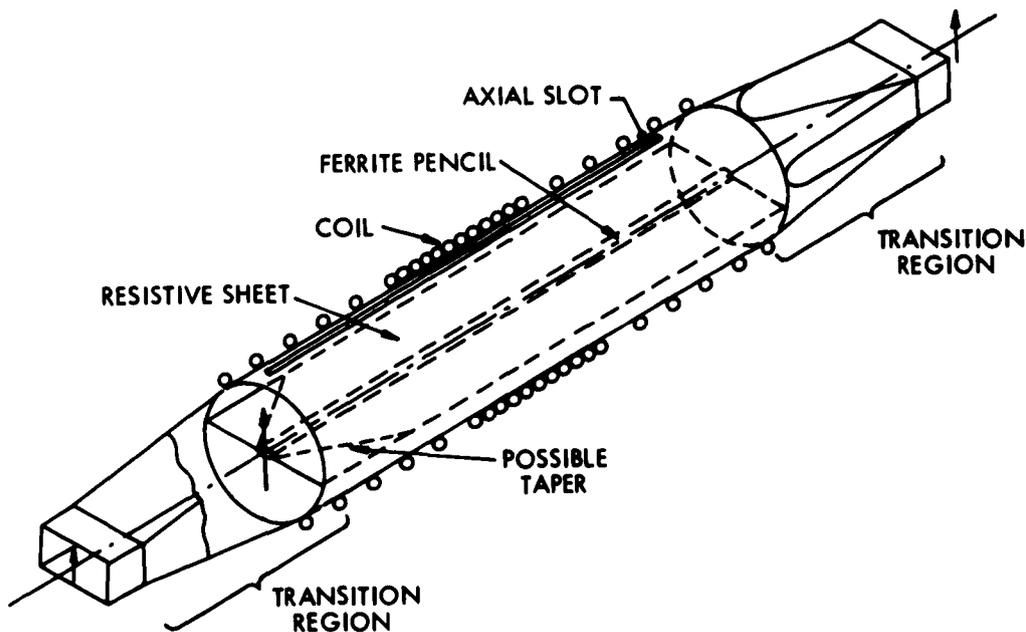
Fig. VII-7. Broadband ferrite switch.

around the axis approximately 0.5-1 turn, so that over the entire band the resistive
sheet attenuates the wave. Low VSWR is obtained by suitably tapering the ferrite pen-
cil, the resistive sheet, the magnetic field, and the waveguide. High switching speed
is obtained by axially slotting the waveguide at the surface current null line.

D. H. Staelin

D. INTERFEROMETERS EMPLOYING CLIPPED SIGNALS

An investigation of a method of eliminating the gain stability requirements for interferometers by processing only the zero crossings of the original signals has been made. If the signals are clipped to a known level, it is possible to re-establish this level at the signal processer regardless of the channel gain. The two-level form created by infinite clipping is in a one-bit quantized form that is suitable for digital signal processing. For some interferometer systems it is desirable to measure the correlation of the two-antenna output signals as a function of antenna spacing and time delay. The correlation process can be performed by a simple coincidence detector when clipped signals are used.

In radio astronomy, the signals most often observed have a Gaussian probability distribution in amplitude. The relationship between the spatial crosscorrelation of the clipped s.gnals and the spatial crosscorrelation of the unclipped signals is similar to the known formula for the temporal autocorrelation of clipped Gaussian signals. If the Gaussian signals have zero means and are symmetrically clipped at $\pm \ell$ volts, then this relationship has the form

$$\phi(\tau, d) = [\phi_{11}(0)\phi_{22}(0)]^{1/2} \sin\left[\frac{\pi}{2\ell^2} \phi_c(\tau, d)\right],$$

where $\phi(\tau, d)$ is the spatial crosscorrelation between the original unclipped signals, $\phi_c(\tau, d)$ is the crosscorrelation between the clipped signals, and $\phi_{11}(0)$ and $\phi_{22}(0)$ are the average powers in channel one and two, respectively, d is antenna spacing, and τ is time delay in one channel.

Once the spatial crosscorrelation has been determined, the angular power density spectrum $P(f, \ell, m)$ may be calculated from a three-dimensional Fourier transform of $\phi(\tau, d)$, where ℓ and m are directional cosines.

As far as it is known, this is the first time that such a technique has been applied to interferometer systems. This new system promises to be very interesting and is one of the first to utilize digital processing of high-frequency interferometric data.

R. K. Breon, J. W. Graham

Prof. K. U. Ingard   A. V. Dralle   T. J. Kinzer III
Dr. L. W. Dean III   S. R. Ellms   F. A. Liégeois
Dr. G. C. Maling, Jr.   S. T. Emerson   W. M. Manheimer
Dr. H. L. Willke, Jr.   J. M. Witting

A. STABILITY OF PARALLEL FLOWS

Heisenberg's discovery of unstable asymptotic solutions for the separated pertur-bation equations for plane viscid Poiseuille flow,[1] together with Rayleigh's proof of the nonexistence of unstable separating solutions in the inviscid case,[2] has encouraged the conjecture that viscous forces can be a cause of instability. When the rarity of sepa-rating solutions among the solutions of the full initial value problem for the perturbed flow is properly taken into account, however, this apparent disconformity with the expected stabilizing role of viscosity disappears. The absence of separating inviscid solutions corresponding to the viscous ones found by Heisenberg is to be viewed as implying not that the inviscid flow is stable, but only that the equations of perturbation fail to separate in the inviscid case.

The equations of perturbation do seperate in general under perturbations that are homogeneous in the direction of the basic flow; otherwise, in the case of inviscid flows without points of inflection, they do not. In the author's thesis,[3] the initial value prob-lem restricted to such perturbations was worked out for an arbitrary basic flow in the inviscid case, and for plane and cylindrical flows in the viscid case. A somewhat less transparent approach was used earlier to discuss the general inviscid and plane viscid cases.[4] It is shown that every inhomogeneous inviscid flow is unstable under such per-turbations. In the viscid case, the solutions follow the corresponding inviscid ones in time until growth by a factor of the order of the Reynolds number is achieved, after which rapid viscosity-controlled decay sets in. Comparison with the exact equations of motion indicates that qualitatively different results are not to be expected when per-turbations of arbitrary amplitude are admitted.[3,4]

In the author's thesis,[5] it is also shown in the plane Couette case, and proposed on physical grounds in the general case, that the inviscid flow is actually not stable under any perturbation at all. Asymptotic separating solutions are found for the viscid Couette case; and these, in contrast with their inviscid counterparts, are all stable. In the absence of boundaries, viscid Couette flow is shown, moreover, to be stable under every sufficiently smooth perturbation.

A brief outline of the work on plane Couette flow[5] is given here.

The fundamental perturbation equation for a plane parallel flow $W(x_1)$ in the x_3 direction has been given[4] in the form

---

$$\frac{\partial}{\partial t}\nabla^2 u_1 = \nu\nabla^4 u_1 - W\frac{\partial}{\partial x_3}\nabla^2 u_1 + \frac{d^2 W}{dx_1^2}\frac{\partial u_1}{\partial x_3}.$$ (1)

For the plane Couette flow $W(x_1) = x_1 V$, Eq. 1 simplifies to

$$\left(\frac{\partial}{\partial t} - \nu\nabla^2 + Vx_1\frac{\partial}{\partial x_3}\right)\nabla^2 u_1 = 0.$$ (2)

The exact free-field propagator for (2) has been found to be

$$K(\vec{x},\vec{y},t) = \frac{1}{(4\pi\nu t)^{3/2}\left(1+\frac{V^2 t^2}{12}\right)^{1/2}}$$

$$\cdot\exp\left\{-\frac{1}{4\nu t}\left[(x_1-y_1)^2 + (x_2-y_2)^2 + \frac{\left(x_3 + Vt\frac{x_1+y_1}{2}-y_3\right)^2}{1+\frac{V^2 t^2}{12}}\right]\right\},$$ (3)

so that, formally, for $t > 0$,

$$\nabla^2 u_1(\vec{x},t) = \iiint K(\vec{x},\vec{y},t)\,\nabla^2 u_1(\vec{y},0)\,d^3 y.$$ (4)

The propagator (3) is similar to the familiar Poisson propagator for the heat equation, the result of the extra terms in (3) being ultimately, if anything, an increase in the damping of the initial disturbance. Thus it is shown readily that in the absence of boundaries the plane viscid Couette flow is stable under every twice-differentiable perturbation $u_1(\vec{x},0)$ that tends to zero, along with its second derivative, for $x^2 \to \infty$. When periodic boundary conditions are imposed in the x_2 and x_3 directions, the same result holds for perturbations that are periodic rather than those that tend to zero in those directions.

In the presence of boundaries, we require, not the free-field propagator, but a propagator maintaining viscous boundary conditions. The term $Vt(x_1+y_1)/2$ in the exponential in (3) rules out the use of the image methods usually employed in constructing bounded-field propagators from free-field ones.

In the absence of a conclusive demonstration of the stability of plane viscid Couette flow in the presence of boundaries, we have sought unstable separating solutions for (2) of the type

$$\nabla^2 u_1(\vec{x},t) = a(x_1)\,e^{st}\,\exp(ik_2 x_2 + ik_3 x_3)$$ (5)

for which (2) reduces to

$$\left(s + \nu k^2 - \nu \frac{d^2}{dx_1^2} + ik_3 Vx_1\right) a(x_1) = 0,$$

(6)

where $k^2 = k_2^2 + k_3^2$.

Equation 6 is a form of Airy's equation, the two basic solutions of which can be written as contour integrals by means of the Laplace method. Accurate asymptotic representations have been found for these integrals through the use of the saddle-point method of integration. The requirement that u_1 vanish, along with its first derivative, at given boundaries, say $x_1 = 0$ and $x_1 = 1$, establishes an eigenvalue problem for s. The problem has solutions, the number and accuracy of which increase without limit as $k_3 V/\nu = k_3 R \to \infty$. All such solutions for $u_1(x,t)$ are stable, and, in fact, decay with time at least as rapidly as

$$\exp\left[-\left(\nu k^2 + \frac{|k_3 V|}{4}\right)t\right].$$

(7)

When $\nu = 0$, the basic equations of perturbation[6] contain no derivatives of the u_i higher than the first, so that $\nabla^2 u_1$ need not exist. When the Laplacian of the initial perturbation $u_1^{(0)}$, as well as $\frac{\partial}{\partial x_3} \nabla^2 u_1$, does exist, however, they will satisfy the inviscid form of (2):

$$\left(\frac{\partial}{\partial t} + Vx_1 \frac{\partial}{\partial x_3}\right)\nabla^2 u_1 = 0.$$

(8)

Under the initial conditions $\nabla^2 u_1^{(0)} = a(x_1, x_2, x_3)$, the relevant solution of (8) will have the familiar form

$$\nabla^2 u_1(x_1, x_2, x_3, t) = a(x_1, x_2, x_3 - Vx_1 t),$$

(9)

which is verifiable by direct substitution.

Equation 9 can be solved for u_1 with the boundary conditions that u_1 vanish at the walls and either be periodic in x_2 and x_3 or tend to zero as x_2 and x_3 become infinite.

In order to use (8) and (9) even when $\nabla^2 u_1$ does not exist, we expand $u_1^{(0)}$ in a Fourier series or Fourier integral in x_2 and x_3, our choice depending upon the boundary conditions. Provided that $\frac{d^2 u_1^{(0)}}{dx_1^2}$ exists, the individual Fourier components[7]

$$u_1^k(x_1, t) \exp(ik_2 x_2 + ik_3 x_3)$$

(10)

for u_1 can be computed from (8) whether or not the full series or integral is twice differentiable with respect to x_2 and x_3. The expansion is then a solution for the first of the basic equations of perturbation,[6] provided that it converges along with its first derivative with respect to x_3.

Substituting (10) in (9), we obtain the equation

$$\left(\frac{\partial}{\partial t} + ik_3 Vx_1\right)(\partial^2 - k^2)\, u_1^k(x_1, t) = 0, \tag{11}$$

where $\partial = \frac{\partial}{\partial x_1}$ and $k^2 = k_2^2 + k_3^2$, so that

$$(\partial^2 - k^2)\, u_1^k(x_1, t) = \exp(-ik_3 Vx_1 t)\, (\partial^2 - k^2)\, u_1^k(x_1, 0). \tag{12}$$

Note that in case $k_3 = 0$, (11) and (12) become trivial, and $u_1^{k_2}(x_1, t) = u_1^{k_2}(x_1, 0)$. This is precisely the case of perturbations that are homogeneous in the direction of flow that was studied in general.[3] The material[5] discussed here thus involves the behavior of the modes with $k_3 \neq 0$.

It can be verified by direct substitution that a particular inversion integral for (12) is

$$u_p^k(x_1, t) = e^{-kx_1} \int_0^{x_1} e^{2kx'} \int_{\epsilon(t)}^{x'} e^{-kx''} \exp(-ik_3 Vx'' t)\, (\partial^2 - k^2)\, u_1^k(x'', 0)\, dx''dx'. \tag{13}$$

Equation 13, as written, vanishes identically for $x_1 = 0$, and it is clear that, for each t, $1 \geqslant \epsilon \geqslant 0$ can be chosen so that (13) will also vanish at $x_1 = 1$ and thus satisfy the boundary conditions for u_1.

By means of estimates of the integrals in (13), certain total integrability conditions on $\frac{\partial^2}{\partial x_1^2} u_1(x_i, 0)$ are derived which suffice for the uniform boundedness of $\sum_k |u_1^k(x_1, t)|$ and therefore of $u_1(x_i, t)$. For example, either

$$\int dx_2 \int dx_3 \left(\int_0^1 \left|\frac{\partial^2}{\partial x_1^2} u_1(x_i, 0)\right| dx_1\right)^2 < \infty$$

or

$$\int dx_2 \int dx_3 \left\|\frac{\partial^2}{\partial x_1^2} u_1(x_i, 0)\right\|_{\infty(x_1)} < \infty \tag{14}$$

suffices.

As a counterexample to much weaker sufficient conditions than those of the character of (14), a perturbation $u_1^*(x_i, 0)$ has been given,[5] which possesses derivatives of all

orders with respect to x_2 and x_3, but only one continuous derivative with respect to x_1, for which the solution $u_1(x_i,t)$ is not bounded in time, but rather oscillates with linearly increasing amplitude for large t.

Under certain conditions $u_1(x_i,t)$ does itself tend to zero for large t, but it does not follow that the flow is stable under these modes of perturbation. The inhomogeneous term $-Vu_1$ in the equation[6] for u_3 yields a contribution to u_3 which fails to vanish for large t even when u_1 does tend to zero, unless $u_1(x_i,0)$ vanishes identically.

The equations of perturbation[6] for u_2 and u_3, for an arbitrary plane parallel flow $W(x_1)$, are

$$\frac{\partial u_i}{\partial t} = -W\frac{\partial u_i}{\partial x_3} - \delta_{i3}\frac{dW}{dx_1}u_1 - \frac{\partial p}{\partial x_i}. \tag{15}$$

One may easily verify that the solution for (15) with given initial conditions $u_i^{(0)}$ can be formally written

$$u_i(x_1,x_2,x_3,t) = u_i^{(0)}(x_1,x_2,x_3-W(x_1)t) - \int_0^t f(x_1,x_2,x_3-W(x_1)(t-t'),t')\,dt' \qquad i = 2, 3, \tag{16}$$

where f represents the inhomogeneous terms on the right-hand side of (15).

It is clear from (16) that the u_i cannot tend uniformly to zero for large t unless $u_2^{(0)}$ and $u_3^{(0)}$ vanish identically. In general, then, we conclude that plane inviscid Couette flow is stable under no nontrivial perturbations whatsoever.

Actually, in view of the absence of apparent means of dissipation, the only surprising element in this result should be the vanishing of u_1 for large t in certain cases. This effect can be "explained" in the following way: as a result of the shearing action of the basic flow, as $t \to \infty$ every neighborhood of a given point (x_1,x_2,x_3) will contain elements of the initial perturbation field from a continuum of points along a tubular neighborhood extending from $(x_1,x_2,-\infty)$ to (x_1,x_2,∞) along the streamlines of the basic flow. One might expect that the vanishing of $u_1(x_i,t)$ at (x_1,x_2,x_3) as $t \to \infty$ could result from the vanishing of the average of the initial perturbation $u_1(x_i,0)$ along streamlines passing near (x_1,x_2,x_3).

The result of the mathematical analysis is essentially the same as that given above. It is clear that, according to the Riemann-Lebesgue theorem, the inner integral in (13), and therefore $u_1^k(x_1,t)$, will vanish for $t \to \infty$, provided that $k_3 \neq 0$ and $\frac{\partial^2}{\partial x_1^2}u_1^k(x_1,0)$ is integrable in x_1.

Because of the uniformity of convergence of $\sum_k |u_1^k(x_1,t)|$, $u_1(x_i,t)$ itself will tend to zero for large t, provided that all of the $u_1^k(x_1,0)$ for $k_3 = 0$ vanish identically. The

latter condition is clearly equivalent to

$$\int_{-\infty}^{\infty} u_1(x_1, 0)\, dx_3 = 0 \qquad \text{for almost all } x_1 \text{ and } x_2,$$

which is the same condition that was stated in words on the basis of the physical argument.

H. L. Willke, Jr.

References

1. W. Heisenberg, Über Stabilität und Turbulenz von Flüssigkeitsströmen, Ann. Phys. **74**, 577-627 (1924).

2. Lord Rayleigh, The Theory of Sound, Vol. 2 (Dover Publications, Inc., New York, 1945), pp. 381-384.

3. H. L. Willke, Jr., Stability Problems in Parallel Flow, Ph. D. Thesis, Department of Physics, M. I. T., June 1963, pp. 17-39.

4. H. L. Willke, Jr., Stability of parallel flows, Quarterly Progress Report No. 68, Research Laboratory of Electronics, M. I. T., January 15, 1963, p. 45.

5. H. L. Willke, Jr., Stability Problems in Parallel Flow, op. cit., pp. 40-77.

6. H. L. Willke, Jr., Stability of parallel flows, op. cit., p. 45, Eq. 1a.

7. It is hoped that no confusion will result from the use of k in superscripts to stand for (k_2, k_3).

B. SOUND EMISSION FROM KARMAN VORTICES

The instability of flow past a cylinder resulting in the familiar Karman vortex street generally represents a very weak sound source. However, we have found that when the cylinder is placed inside a tube and perpendicular to the axis, an intense oscillation is produced whenever the frequency of the Karman vortex shedding coincides with one of the cross-mode resonances in the tube. At these frequencies the standing wave established across a tube stimulates the instability of the flow in the tube, and through this feedback mechanism a self-sustained fluid oscillation of considerable magnitude results. The frequency of shedding of the Karman vortices is known empirically to be $f_o = aV/d$, where V is the flow velocity and d the cylinder diameter. The Strouhal number a is a weak function of Reynolds number R and is approximately a constant $a \approx 0.18$ in the region $10^2 < R < 10^5$. In the case of a rectangular tube, the relevant n^{th} cross-mode resonance occurs at a wavelength $\lambda_n = 2D/n$, where D is the transverse horizontal dimension of the tube, which is perpendicular to the cylinder. Strong coupling between the Karman vortex street and the transverse acoustic modes, then, occurs when $f_o = f_n = nc/2D$, where c is the speed of sound and n an integer. Using the value for f_o given above, we see that the condition for resonance can also be written

$$V = \left(\frac{n}{2a}\right)\frac{d}{D} c \approx 2.8n\frac{d}{D} c.$$

Although in free space the spectrum of the sound from the Karman vortices contains practically only the fundamental frequency f_o, the resonance oscillations in the tube contain a large number of harmonics.

The feedback between the sound field and the vortex street is produced by the velocity field in the sound field. This has been demonstrated by measuring the intensity of the sound field as a function of position of the cylinder in the direction perpendicular to the tube axis. Maximum intensity is obtained when the cylinder is placed in a maximum velocity of the sound field.

A similar feedback oscillator involving two interacting cylindrical rods has also been studied. When two rods are placed one after the other and perpendicular to the flow, a resonance oscillation occurs when the time of convection of a vortex from one cylinder to the other is equal to the period of the Karman tone. Since the drift velocity of a vortex is approximately 80 per cent of the mean flow speed in the tube, it follows that the condition for such a resonance is simply $0.8\frac{L}{V} = \frac{1}{f_o}$, which corresponds to a separation of the cylinders which is equal to approximately four times the cylinder diameter.

A more detailed account of these studies is being prepared for publication.

K. U. Ingard, W. M. Manheimer

C. GENERATION OF SHOCK WAVES BY MEANS OF EXPLODING METALLIC FILMS

Shock waves produced by discharging a condenser bank through film of aluminized Mylar have been studied in air at various pressures. High-frequency electrostatic transducers have been used as probes for the determination of shock speed and the measurement of shock strength and reflection coefficients. A preliminary analysis of the data indicates that the shock speed is proportional to the square root of discharged energy and, at least at low pressures, proportional to the fourth power of the ambient density. Also at low pressures electrical effects attributed to plasma flow behind the shock were observed. A detailed analysis of the data is now under way.

S. T. Emerson, A. V. Dralle, K. U. Ingard

IX. NOISE IN ELECTRON DEVICES

Prof. H. A. Haus
Prof. P. L. Penfield, Jr.

Prof. R. P. Rafuse
Dr. G. Fiocco

A. HIGHER-ORDER CORRELATION FUNCTIONS OF LIGHT INTENSITY

A measurement of the photoelectron count of a photosurface illuminated by laser light was outlined in Quarterly Progress Report No. 69 (pages 31-33). It was pointed out there that the mean-square deviation of the photoelectron count within a time T is related to the correlation function of the light intensity $R(T)$.[1,2] We shall now show that higher-order moments of the photoelectron count are related to the higher-order correlation functions of the light intensity. In the derivation we assume that the coherence area of the light beam is larger than the photocathode area.

The probability $p_T(K)$ of obtaining exactly K counts in the time interval $0 \leqslant t \leqslant T$ from a photocathode illuminated by light of intensity (power) $P(t)$ is

$$p_T(K) = \frac{n_T^K}{K!} e^{-n_T}, \tag{1}$$

where

$$n_T = \alpha \int_0^T P(t)\, dt. \tag{2}$$

Here, α is a proportionality factor incorporating the photoefficiency of the cathode. The falling factorial moment of kth order of the photoelectron count, $\overline{K(K-1) \ldots (K-k+1)}$, is

$$\overline{K(K-1) \ldots (K-k+1)} = \sum_{K=1}^{\infty} \frac{n_T^K}{K!} e^{-n_T} K(K-1) \ldots (K-k+1) = n_T^k. \tag{3}$$

Thus far, we have taken an average with respect to the probability distribution of the Poisson process. If the light intensity itself varies in a statistical manner, an average has to be taken over the statistics of the incident light. This average is denoted by angular brackets.

$$\left\langle \overline{K(K-1) \ldots (K-k+1)} \right\rangle = \left\langle n_T^k \right\rangle = \alpha^k \left\langle \left[\int_0^T P(t)\, dt \right]^k \right\rangle. \tag{4}$$

The right-hand side of Eq. 4 may be written

$$\left\langle n_T^k \right\rangle = k!\, \alpha^k \int_{t_k=0}^T \ldots \int_{t_1=0}^{t_2} \left\langle P(t_1) \ldots P(t_k) \right\rangle dt_1 \ldots dt_k. \tag{5}$$

The integrand can be put into the form of a $(k-1)^{th}$-order correlation function

by introducing the new variables

$$t_1 = t; \ t_2 = t + \tau_1; \ \ldots \ t_k = t + \tau_{k-1} \tag{6}$$

$$\left\langle n_T^k \right\rangle = k! \, a^k \int_{\tau_{k-1}=0}^{T} \int_{t=0}^{T-\tau_{k-1}} \ldots \int_{\tau_1=0}^{\tau_2} \left\langle P(t) \ldots P(t+\tau_{k-1}) \right\rangle d\tau_1 \ldots dt d\tau_{k-1}.$$

If the light is stationary, the expectation value of the integrand does not depend upon t, and the integration over t can be carried out immediately. Furthermore, noting that the integrand is the correlation function of $(k-1)^{th}$ order, $R_{k-1}(\tau_1 \ldots \tau_{k-1})$, of the light power, one obtains

$$\left\langle n_T^k \right\rangle = k! \, a^k \int_{\tau_{k-1}=0}^{T} \ldots \int_{\tau_1=0}^{\tau_2} (T-\tau_{k-1}) \, R_{k-1}(\tau_1, \ldots \tau_{k-1}) \, d\tau_1 \ldots d\tau_{k-1}. \tag{7}$$

Differentiating Eq. 7 with respect to T, one obtains

$$\frac{d}{dT} \left\langle n_T^k \right\rangle = k! \, a^k \int_{\tau_{k-1}=0}^{T} \ldots \int_{\tau_1=0}^{\tau_2} R_{k-1}(\tau_1, \ldots \tau_{k-1}) \, d\tau_1 \ldots d\tau_{k-1}. \tag{8}$$

The falling factorial moment of kth order of the photoelectron count within an observation time T is thus simply related to the $(k-1)^{th}$-order correlation function of the light intensity. For the factorial moment of second order, $k = 2$, one obtains

$$\frac{d^2}{dT^2} \left\langle n_T^2 \right\rangle = 2a^2 R_1(T). \tag{9}$$

H. A. Haus

References

1. E. M. Purcell, The question of correlation between photons in coherent light rays, Nature **178**, 1449 (1956).

2. L. Mandel, Fluctuations of photon beams and their correlations, Proc. Roy. Soc. (London) **72**, 1037-1048 (1958).

X. ELECTRODYNAMICS OF MOVING MEDIA

Prof. L. J. Chu
Prof. H. A. Haus
Prof. P. Penfield, Jr.

A. RESEARCH OBJECTIVES

Finding the force of electromagnetic origin in a continuous medium is an old problem that has not been solved except in rather special cases. (An exception is the work of Meixner and his associates.[1-4]) It involves the unification of two physical theories — continuum mechanics and electromagnetism. Since nonrelativistic mechanics is not, in general, compatible with electromagnetism, relativistic mechanics must be used.

The relativistic theory of continuum mechanics (without electromagnetic fields) is fairly well understood, at least insofar as its basic equations are concerned. Electromagnetism in stationary bodies is also fairly well understood.

To unite the two theories, the effect of each upon the other must be calculated. The study of the effect of mechanical motion and deformation upon electromagnetism is known as "electromagnetism of moving media," and there are several theories that are formally different but lead to the same physical predictions.[5] The inverse problem, finding the effects of the electromagnetic fields upon the mechanical equations, has never been solved adequately, and it is this problem that we wish to discuss now.

The fundamental equations of relativistic continuum mechanics are: (1) Newton's Law, which, in its simplest form, is expressed as

$$n \frac{d}{dt}(m\bar{v}) = \bar{f}_{mech},$$ (1)

where n is the number of particles per unit volume, m the relativistic mass per particle, $\bar{v}$ the velocity, and $\bar{f}_{mech}$ the force density, now of entirely mechanical origin; and (2) a specification of the way in which $\bar{f}_{mech}$ is related to n, the strain, the strain rate or other kinematic variables. That is, the fundamental equations are Newton's Law, and the mechanical (and thermodynamic) constitutive relations of the material.

The fundamental equations of electromagnetism of stationary bodies are Maxwell's equations, and relations among the field quantities, that is, electromagnetic constitutive relations of the material.

It is instructive to consider how mechanical motion affects electromagnetism because the inverse problem, broadly speaking, is similar. Fundamentally, mechanical motion can affect electromagnetism in two ways. First, additional terms may enter, as source terms, into Maxwell's equations. Second, the constitutive relations may be altered in form.

There is essentially only one theory of electromagnetism in free space, but, if materials are present, there are several, including (among others) the Minkowski

theory, the theory of L. J. Chu, and the Amperian current-loop theory.[5] These different theories are possible because macroscopic electric and magnetic fields cannot be measured inside material, but must be inferred from measurements performed nearby in free space. The various theories differ with respect to the manner in which this inference is made. They disagree with respect to the field variables used, the form of Maxwell's equations, and the form of the constitutive laws, especially when the material is moving and deforming. Thus, in Minkowski's theory, Maxwell's equations have the same form in free space, in stationary material, and in moving material. However, the electromagnetic constitutive relations are altered by the material and by the motion and deformation. On the other hand, in the Chu and Amperian theories, source terms are introduced into Maxwell's equations if stationary material is present, and these are modified further if there is motion or deformation. Constitutive relations are not necessary in free space, but are required for both stationary and moving media.

In spite of their differences, these theories are all equivalent,[5] in the sense that they apply to the same physical situations and make identical predictions of fields outside material bodies. Since fields cannot be measured inside material, the theories cannot be distinguished by any purely electromagnetic measurements.

Now consider the inverse problem, the effect of electromagnetic fields on the mechanical equations. This effect can consist of two parts: First, Newton's law will include an additional force term, the "force of electromagnetic origin,"

$$n \frac{d}{dt}(m\bar{v}) = \bar{f}_{mech} + \bar{f}_{em}. \tag{2}$$

Second, the mechanical constitutive laws may be altered by the fields, so that $\bar{f}_{mech}$ depends upon both mechanical and electromagnetic variables. It should be clear that there is no unique way of separating these two effects. Since the ultimate purpose is to solve Eq. 2, only the sum $\bar{f}_{mech} + \bar{f}_{em}$ has significance, and additional terms may be considered as a part either of the first or second term. The viewpoint that we shall use, but which is not universal, is to retain the mechanical constitutive relations unchanged in form, so that $\bar{f}_{mech}$ can be computed in the normal way from the mechanical variables, and does not depend explicitly on the electromagnetic fields. Thus all additional forces are to be considered as part of $\bar{f}_{em}$.

One reason why this problem has not been solved previously is that many workers[6-11] derived a force of electromagnetic origin (or, the equivalent, an electromagnetic stress-energy tensor), but did not ask how the mechanical constitutive relations are altered by the field. The various forces or stress-energy tensors that were obtained did not agree with each other, and there has been considerable discussion[7, 12] about the "correct" force or tensor. What has been said, thus far, indicates that such discussions are irrelevant unless the particular force or tensor is accompanied by a

statement of changes to be made in the mechanical constitutive relations.

On the other hand, some workers[1-4, 13-17] have appreciated this point, and have usually accounted for it by giving the total force, or the total stress-energy tensor, that is, the sum of the mechanical and electromagnetic parts. Of these workers, only Meixner and his associates[1-4] have obtained results that agree with ours in the common area of application. Meixner has treated the thermodynamic aspects more thoroughly than we have, but has considered only linear rest-frame electrical constitutive relations. We allow for nonlinear relations, partly for greater generality, but also for easier physical interpretation of many terms. We believe that our derivations of the force density use fewer, more reasonable postulates than that of Meixner and de Sa,[1] and also require somewhat less complicated mathematics. Furthermore, we wanted to, and were able to, reconcile the apparently different force densities predicted by the various theories of electromagnetism. This is important because some of the theories (especially the Chu and Amperian theories) have rather simple physical interpretations in terms of microscopic models, and we are now able to extend these interpretations to the force density.

In Section X-B, the relativistic force of electromagnetic origin is derived by use of Hamilton's Principle.

<div align="right">L. J. Chu, H. A. Haus, P. Penfield, Jr.</div>

References

1. B. de Sa, Der Energie-Impuls-Tensor in der Relativistischen Thermodynamik Irreversibler Prozesse, Dr. Nat. Thesis, Technische Hochschule, Aachen, Germany; February 9, 1960.

2. C. L. Tang and J. Meixner, Relativistic theory of the propagation of plane electromagnetic waves, Phys. Fluids 4, 148-154 (1961).

3. J. Meixner, Relativistic Thermodynamics of Irreversible Processes in a One Component Fluid in the Presence of Electromagnetic Fields, Repc t RL-184, Radiation Laboratory, Department of Electrical Engineering, University of . ichigan, April 1961 (unpublished).

4. J. Meixner, Der Drehimpulssatz in der Thermodynamik der irreversiblen Prozesse, Z. Physik 164, 145-155 (1961).

5. R. M. Fano, L. J. Chu, and R. B. Adler, Electromagnetic Fields, Energy, and Forces (John Wiley and Sons, Inc., New York, 1960), pp. 453-486.

6. M. Abraham and R. Becker, The Classical Theory of Electricity and Magnetism (Blackie and Son Ltd., London, 1937).

7. C. Møller, The Theory of Relativity (Oxford University Press, London, 1952), pp. 202-211.

8. J. L. Synge, Relativity: The Special Theory (North-Holland Publishing Company, Amsterdam, and Interscience Publishers, Inc., New York, 1956).

9. L. V. Boffi, Electrodynamics of Moving Media, Sc.D. Thesis, Department of Electrical Engineering, M.I.T., August 19, 1957.

10. N. H. Meyers, A Poynting Theorem for Moving Bodies and the Relativistic Mechanics of Extended Objects, J. Franklin Inst. 266, 439-464 (December 1958).

11. R. M. Fano, L. J. Chu, and R. B. Adler, op. cit., pp. 487-499.

12. W. Pauli, Theory of Relativity (Pergamon Press, New York, 1958), pp. 99-111.

13. G. Marx and G. Györgyi, Der Energie-Impuls-Tensor des Elektromagnetischen Feldes und die Ponderomotorischen Kräfte in Dielektrika, Acta Phys. Acad. Sci. Hungaricae, Vol. 3, pp. 213-242, 1954.

14. G. Marx, Das Elektromagnetische Feld in Bewegten Anisotropen Medien, Acta Phys. Acad. Sci. Hungaricae, Vol. 3, pp. 75-94, 1954.

15. G. Marx and K. Nagy, Der Energie-Impuls-Tensor der Strahlung in Dielektrika, Acta Phys. Acad. Sci. Hungaricae, Vol. 4, pp. 297-300, 1955.

16. G. Györgyi, Die Bewegung des Energiemittelpunktes und der Energie-Impuls-Tensor des Elektromagnetischen Feldes in Dielektrika, Acta Phys. Acad. Sci. Hungaricae, Vol. 4, pp. 121-131, 1955.

17. Pham Mau Quan, Étude électromagnétique et thermodynamique d'un fluide relativiste chargé, J. Rational Mech. Anal. 5, 473-538 (1956).

B. FORCE OF ELECTROMAGNETIC ORIGIN IN FLUIDS

Here, we derive the force of electromagnetic origin on a polarizable and magnetizable fluid, using Hamilton's principle as applied to three common theories of electromagnetism of moving media. Hamilton's principle for a dielectric and magnetic fluid is an extension of that for an ordinary fluid because terms that account for the electromagnetic fields and the electromagnetic properties of the material are added to the Lagrangian for ordinary fluids. We shall discuss (a) the fluid without electromagnetic fields, (b) a stationary electromagnetic medium, and (c) the electromagnetic fluid.

1. Hamilton's Principle for a Fluid without Electromagnetic Fields

Until recently, only irrotational flow was predicted by Hamilton's principle for an ordinary fluid, but Lin[1-3] has shown how an additional constraint on the variations leads to rotational flow also. The Lagrangian density, in Eulerian coordinates, is

$$L = W_k^i - W_{f0}. \tag{1}$$

Here,

$$W_k^i = -n_0 m_0 c^2 \tag{2}$$

is the relativistic kinetic co-energy per unit volume, with n_0 the particle density in the rest frame, m_0 the rest mass per particle, and W_{f0} the rest-frame fluid intrinsic energy or free energy per unit volume. We must appeal to thermodynamics to find the form for W_{f0}. For example, for an electrically neutral fluid with an isothermal relation[4] between density n_0 and pressure p_f in the rest frame, W_{f0} is a function

of only n_0:

$$W_{f0} = n_0 \int \frac{p_f}{n_0} \, dn_0. \tag{3}$$

However, the laboratory-frame density n is, relativistically, different from n_0

$$n = \gamma n_0, \tag{4}$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}, \tag{5}$$

and hence W_{f0} must be considered a function of density n and (through γ) the velocity $\overline{v}$. Note that W_{f0} could be, as in this case, the free energy (because the isothermal relation between n_0 and p_f is used), or the energy (if the adiabatic relation were used).

The Lagrangian density L is now a function of n and $\overline{v}$, and the first-order variation in $L = L_k + L_f$ is of the form

$$\delta L = (U_k + U_f) \, \delta n + (\overline{V}_k + \overline{V}_f) \cdot \delta \overline{v}, \tag{6}$$

where the U and $\overline{V}$ quantities are determined by taking derivatives of L. The dimensions of U and $\overline{V}$ are energy and momentum per unit volume, respectively.

However, n and $\overline{v}$ cannot be varied independently because of two constraints. One is the law of conservation of particles (continuity equation):

$$\nabla \cdot (n\overline{v}) + \frac{\partial n}{\partial t} = 0. \tag{7}$$

The other is Lin's constraint.[1-3] When these two constraints are used and the corresponding Lagrange multipliers eliminated, Hamilton's principle yields an equation that relates the U and $\overline{V}$ parameters:

$$n\nabla U = n \frac{d}{dt} \frac{\overline{V}}{n} + (\overline{v} \cdot \nabla)\overline{v} + \overline{v} \times (\nabla \times \overline{v}), \tag{8}$$

where $U = U_k + U_f$ and $\overline{V} = \overline{V}_k + \overline{V}_f$.

Substituting

$$U_k = -m_0 c^2 / \gamma \tag{9}$$

$$\overline{V}_k = nm\overline{v}, \tag{10}$$

which are derived from the known form of W_k', and in which $m = \gamma m_0$ is the relativistic mass, we find

$$n \frac{d}{dt} (m\overline{v}) = n\nabla U_f - n \frac{d}{dt} \frac{\overline{V}_f}{n} - (\overline{V}_f \cdot \nabla)\overline{v} - \overline{V}_f \times (\nabla \times \overline{v}).$$

It is now clear that the right-hand side of this equation is the force density of mechanical origin. In the case of the particular constitutive relation leading to Eq. 3,

$$U_f = -\frac{W_{f0} + P_f}{n},$$
(12)

$$\overline{V}_f = \frac{\gamma^2 \overline{v}}{c^2}(W_{f0} + P_f),$$
(13)

and hence the force density of mechanical origin is the well-known result[5]

$$-\nabla P_f - n\frac{d}{dt}\frac{\overline{V}_f}{n},$$
(14)

where $\overline{V}_f$, a relativistic quantity, is the momentum density associated with the rest energy and the rest-frame stress.

2. Hamil·on's Principle for Electromagnetism of Stationary Media

The Lagrangian for an electromagnetic field is of the form

$$L = W'_e - W_m.$$
(15)

where W'_e is some electric co-energy and W_m is a magnetic energy. The particular forms of W'_e and W_m depend both upon the theory of electromagnetism that is used and upon the constitutive laws of the material. Ultimately, we must appeal to thermodynamics to tell us the form of the energy densities (or, if isothermal constitutive relations are used, the free-energy densities). From the electric energy density, we perform a Legendre transformation to get W'_e.

Consider now a stationary dielectric and/or magnetic material. We shall use Chu's theory of electromagnetism.[6] We know from thermodynamics that the electric energy density is

$$\int \overline{E} \cdot d(\epsilon_0\overline{E} + \overline{P}),$$
(16)

which breaks apart naturally into a portion resulting from the field and a portion resulting from the material. Subtracting Eq. 16 from $\overline{E} \cdot (\epsilon_0\overline{E} + \overline{P})$, we find that

$$W'_e = \frac{1}{2}\epsilon_0 E^2 + \int_0^{\overline{E}} \overline{P} \cdot dE.$$
(17)

The magnetic energy density is, similarly,

$$W_m = \frac{1}{2}\mu_0 H^2 + \int_0^{\mu_0\overline{M}} \overline{H} \cdot d(\mu_0\overline{M}). \tag{18}$$

The Lagrangian is a function of $\overline{E}$, $\overline{H}$, and $\overline{M}$. These three variables are not all independent, however, but are related by two of Maxwell's equations

$$\nabla \times \overline{E} = -\frac{\partial}{\partial t}(\mu_0\overline{H}+\mu_0\overline{M}) \tag{19}$$

$$\nabla \cdot (\mu_0\overline{H}+\mu_0\overline{M}) = 0. \tag{20}$$

If arbitrary variations of L are made subject to these two constraints, Hamilton's principle yields the other two of Maxwell's equations

$$\nabla \times H = \frac{\partial}{\partial t}(\epsilon_0\overline{E}+\overline{P}) \tag{21}$$

$$\nabla \cdot (\epsilon_0\overline{E}+\overline{P}) = 0. \tag{22}$$

Similar derivations can be made by using the other theories of electromagnetism.[7]

3. Hamilton's Principle for Moving Fluids with Electromagnetic Fields

The Lagrangian density for a moving, deforming fluid is found by adding the Lagrangian for the fluid without electromagnetic fields to the electromagnetic Lagrangian. The portions of the Lagrangian in section 2 that deal with the fields ($\overline{E}$ and $\overline{H}$) alone can be written in the same form, but the portions involving $\overline{P}$ and $\overline{M}$ must be evaluated in the rest frame of the material. Furthermore, the electrical constitutive relations may depend upon the density of the fluid; thus, for example, the relation between $\overline{P}_0$ and $\overline{E}_0$ becomes

$$\overline{P}_0 = \overline{P}_0(\overline{E}_0, n_0), \tag{23}$$

where the subscript zero indicates evaluation in the rest frame. The polarization energy density is

$$W_{p0} = \int \overline{E}_0 \cdot d\overline{P}_0 \tag{24}$$

and the polarization co-energy density is

$$W'_{p0} = \int \overline{P}_0 \cdot d\overline{E}_0, \tag{25}$$

where the integration is performed at the actual value of n_0. We define π_p, the

'polarization pressure," as

$$\pi_p = n_0 \frac{\partial W_{p0}}{\partial n_0} + \overline{E}_0 \cdot \overline{P}_0 - W_{p0}. \tag{26}$$

Let us consider a fluid first without magnetization and then with magnetization. The over-all Lagrangian density is

$$L = W_k' - W_{f0} + W_{p0}' + \frac{1}{2} \epsilon_0 E^2 - \frac{1}{2} \mu_0 H^2,$$

a function of n, $\overline{v}$, $\overline{E}$, and $\overline{H}$. Thus the first-order variation of L is of the form

$$\delta L = (U_k + U_f + U_p)\delta n + (\overline{V}_k + \overline{V}_f + \overline{V}_p) \cdot \delta \overline{v} + \overline{W}_p \cdot \delta \overline{E} + \overline{X}_p \cdot \delta \overline{H}, \tag{27}$$

where U_k, U_f, $\overline{V}_k$, and $\overline{V}_f$ are as given above. The last two terms in Eq. 27 lead to Maxwell's equations, including polarization charge and current terms, and the force of electrical origin is found from U_p and $\overline{V}_p$ according to a formula just like that for the force of mechanical origin:

$$\overline{f}_{em} = n\nabla U_p - n\frac{d}{dt}\frac{\overline{V}_p}{n} - (\overline{V}_p \cdot \nabla)v - \overline{V}_p \times (\nabla \times \overline{v}). \tag{28}$$

In our example,

$$U_p = -\frac{W_{p0} + \pi_p - \overline{E}_0 \cdot \overline{P}_0}{n} \tag{29}$$

$$\overline{V}_p = \overline{G}_p - \overline{P} \times \mu_0 \overline{H}, \tag{30}$$

where $\overline{G}_p$ is the (relativistic) momentum associated with the rest-frame energy and stress,

$$\overline{G}_p = \frac{\gamma^2 \overline{v}}{c^2}(W_{p0} + \pi_p) - \frac{\gamma^2 (\overline{P} \cdot \overline{v})}{c^2}(\overline{E} + \overline{v} \times \mu_0 \overline{H}). \tag{31}$$

The force density of electrical origin is

$$\overline{f}_{em} = (\overline{P} \cdot \nabla)\overline{E} + n\frac{d}{dt}\left(\frac{\overline{P}}{n}\right) \times \mu_0 \overline{H} + \nabla \times (\overline{P} \cdot \nabla)\mu_0 \overline{H} - \nabla \pi_p - n\frac{d}{dt}\frac{\overline{G}_p}{n}. \tag{32}$$

This force density has an interesting physical interpretation. Aside from the terms involving π_p and $\overline{G}_p$, it is equal to n times the force that would be exerted on a pair of charges with dipole moment $\overline{P}/n$ in arbitrary electric and magnetic fields.

Forces resulting from magnetization are of a similar form. If these, and the force

density $\rho_{free}\overline{E} + \overline{J}_{free} \times \mu_0\overline{H}$ from free charge and current are added, the over-all force of electromagnetic origin is given bv

$$\overline{f}_{em} = \rho_{free}\overline{E} + \overline{J}_{free} \times \mu_0\overline{H} - \nabla(\pi_p + \pi_m) + (\overline{P} \cdot \nabla)\overline{E} + (\mu_0\overline{M} \cdot \nabla)\overline{H}$$

$$+ \overline{v} \times (\overline{P} \cdot \nabla)\mu_0\overline{H} - \overline{v} \times (\mu_0\overline{M} \cdot \nabla)\epsilon_0\overline{E} + n\frac{d}{dt}\left(\frac{P}{n}\right) \times \mu_0\overline{H}$$

$$+ n\frac{d}{dt}\left(\frac{\mu_0\overline{M}}{n}\right) \times \epsilon_0\overline{E} - n\frac{d}{dt}\left(\frac{\overline{G}_p + \overline{G}_m}{n}\right). \tag{33}$$

The same technique can also be used with the Amperian or the Minkowski theory of electromagnetism. For example, in the Minkowski theory, the rest-frame values of $\overline{E}$ and $\overline{D}$ are related by

$$\overline{E}_0 = \overline{E}_0(\overline{D}_0, n_0), \tag{34}$$

and thermodynamics tells us that the free-energy density (or, if adiabatic constitutive relations are used, the energy density) has as its electrical component

$$W_e = \int \overline{E}_0 \cdot d\overline{D}_0. \tag{35}$$

If we similarly form the magnetic energy and find the electric co-energy, the over-all Lagrangian is of the form

$$L = W'_k - W_{f0} + \int \overline{D}_0 \cdot d\overline{E}_0 - \int \overline{H}_0 \cdot d\overline{B}_0. \tag{36}$$

This is a function of n, $\overline{v}$, $\overline{E}$, and $\overline{B}$. The force of electromagnetic origin derived from it is numerically equal to Eq. 33, although in quite different form, since the field variables have different meanings in the two theories of electromagnetism.

Similarly, using the Amperian formulation of electromagnetism, we start with a Lagrangian density

$$L = W'_k - W_{f0} + \frac{1}{2}\epsilon_0 E^2 - \frac{B^2}{2\mu_0} + \int_0^{\overline{E}_0} \overline{P}_0 \cdot d\overline{E}_0 + \int_0^{\overline{B}_0} \overline{M}_0 \cdot d\overline{B}_0, \tag{37}$$

which is a function of n, $\overline{v}$, $\overline{E}$, and $\overline{B}$. The force of electromagnetic origin derived from it is numerically equal to Eq. 33, although in a different form because the field variables now have meanings different from those in the Chu theory.

P. Penfield, Jr.

References

1. C.-C. Lin, Private Communication, 1959.

2. J. Serrin, Mathematical principles of classical fluid mechanics, Handbuch der Physik, Vol. VIII/1 (Fluid Dynamics I.), edited by S. Flügge (Springer-Verlag, Berlin, 1959), pp. 125-263; see especially p. 148.

3. P. Penfield, Jr., Frequency-Power Formulas (The Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, N.Y., 1960), Section 6.6.

4. H. Lamb, Hydrodynamics (Dover Publications, Inc., New York, Sixth edition, 1945), Article 9.

5. L. D. Landau and E. M. Lifshitz, Fluid Mechanics (Pergamon Press, London and Addison-Wesley Publishing Company, Inc., South Reading, Mass., 1959), Chapter 15.

6. R. M. Fano, L. J. Chu, and R. B. Adler, Electromagnetic Fields, Energy, and Forces (John Wiley and Sons, Inc., New York, 1960).

7. Ibid., pp. 481-486.

PLASMA DYNAMICS

XI. PLASMA PHYSICS*

Prof. S. C. Brown
Prof. G. Bekefi
Prof. D. R. Whitehouse
M. L. Andrews
V. Arunasalam
C. D. Buntschuh
J. F. Clarke
J. D. Coccoli

F. X. Crist
E. W. Fitzgerald, Jr.
W. H. Glenn, Jr.
E. B. Hooper, Jr.
J. C. Ingraham
P. W. Jameson
R. L. Kronquist

D. T. Llewellyn-Jones
J. J. McCarthy
W. J. Mulligan
J. J. Nolan, Jr.
H. R. Radoski
G. L. Rogoff
F. Y-F. Tse
R. E. Whitney

A. EXCITATION OF THE OPTICAL SPECTRUM IN THE HOLLOW-CATHODE ARC

1. Introduction

The intensity of a spectral line from a unit volume of plasma, with self-absorption and broadening neglected, is given by

$$I_{jk} = n_j A_{jk} h\nu_{jk},$$ (1)

where n_j is the density of emitting atoms or ions in the jth excited state, A_{jk} is the transition probability from the jth state to the kth state, and $h\nu_{jk}$ is the photon energy. In a low-pressure arc plasma, which is far from thermal equilibrium, it is necessary to consider the rates of the various populating processes in order to arrive at an expression for n_j in terms of the plasma properties.

The plasma that we are studying is a highly ionized argon arc.[1] The base pressure before ionization is a few μ Hg. The electron density is of the order of 10^{13} cm^{-3}; electron temperatures are several volts at the center of the column. Ion temperatures may be one volt, or more.[2] The spectrum from 3200 Å to 5200 Å contains, for the most part, AII lines and several weak AI and AIII lines.

In a plasma the population process is frequently predominantly electronic excitation from the ground state, and the depopulation process is principally spontaneous emission to all lower states. In this case,

$$n_j = \frac{\beta_{oj} n_o n_-}{A_j},$$ (2)

where β_{oj} is the excitation coefficient, n_o is the density of atoms or ions in the ground state, and A_j is the transition probability per second to all lower states. We shall use

this expression in analyzing the data; however, for the conditions obtaining in the arc, radiative recombination, stepwise electronic excitation, and cascading may give significant contributions to n_j. In this report we consider these other processes and estimate their rates relative to electronic excitation. We shall find that the competing processes may have rates comparable to, but somewhat smaller than, electronic excitation, so that Eq. 2 represents a reasonably good approximation.

In plasmas at micron pressures there is negligible self-absorption of nonresonance radiation at optical frequencies, and negligible population by collisions of the second kind. We shall also neglect the effects of metastables. This is probably not legitimate for the argon atom. If the ion has any metastable levels, they are probably close enough to the radiating levels (less than T_+) to be considered as merged into a single radiating state. Throughout this report a Maxwellian electron distribution is assumed. This is reasonable in view of the high electron temperatures and high degree of ionization. Because none of the relevant cross sections and atomic constants are available for the argon ion, and very few for the atom, it has been necessary to make order-of-magnitude guesses of the values.

2. Excitation Coefficients

The coefficient β_{ij} for excitation from the ith to the jth state by electron impact is the average over the electron distribution function of the cross section σ_{ij} multiplied by the electron velocity v. For a Maxwellian at temperature T_- volts,

$$\beta_{ij} = \sqrt{\frac{8eT_-}{\pi m}} \int_{V_x/T_-}^{-\infty} \sigma_{ij}(u) \frac{u}{T_-} e^{-u/T_-} d\frac{u}{T_-}, \tag{3}$$

where u is the electron energy, and $V_x = V_j - V_i$ is the difference in excitation potential. The ionization coefficient $\beta_{o\infty}$ is also given by Eq. 3.

There are very limited data on the cross sections, especially for argon. There are no data for the excitation or ionization from the ground state of the ion, or from higher levels of either the atom or ion. We can obtain a useful approximation by noting that all of the cross sections have the general form shown by the solid line in Fig. XI-1. For excitation, σ_m is of the order of 10^{-19} cm^2, and V_m is slightly greater than V_x. For ionization, $\sigma_m \sim 10^{-16}$ cm^2, and V_m is several times V_∞, the ionization potential. Three approximations to $\sigma(u)$ which permit Eq. 3 to be readily integrated are shown as dotted lines:

$$\left.\begin{array}{l} \sigma_1 = c_1/u \\ \sigma_2 = c_2(u-V_x) \\ \sigma_3 = \sigma_m \end{array}\right\} \quad u > V_x$$

ELECTRON ENERGY
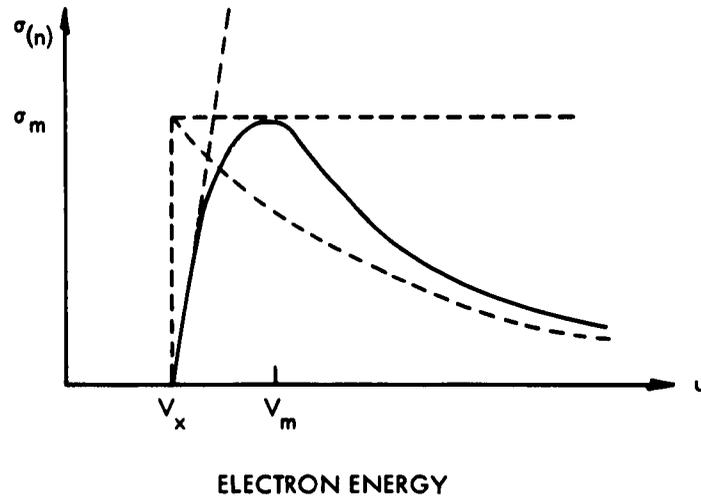
Fig. XI-1. Approximate excitation and ionization cross sections.

For σ_1, we use the theoretical formula $\sigma = C(\ln u)/u$, suppressing the slowly varying logarithm.[3] σ_2 is the popular approximation for low T; σ_3 improves on this for higher T. The resulting coefficients, in mks units, are

$$\beta_1 = C_1 \sqrt{\frac{8e}{\pi m}} T_-^{-1/2} e^{-V_x/T_-}$$

$$\beta_2 = C_2 V_x \sqrt{\frac{8e}{\pi m}} \left(1 + \frac{2T_-}{V_x}\right) T_-^{-1/2} e^{-V_x/T_-}$$

$$\beta_3 = \sigma_m V_x \sqrt{\frac{8e}{\pi m}} \left(1 + \frac{T_-}{V_x}\right) T_-^{-1/2} e^{-V_x/T_-}.$$

At low electron temperatures the exponent contains the principal variation of β with T_-, and it makes little difference which coefficient is used. At high temperatures, the first expression, which takes into account the decreasing cross section at high energy, is presumably the best approximation. For this reason, we use β_1, choosing $C_1 = \sigma_m V_\infty$.

For argon atoms, $V_\infty = 15.8$ volts, we have

$$\beta_{ij}^o = 1.1 \times 10^{-10} T^{-1/2} e^{-V_x/T} \text{ cm}^3/\text{sec}. \tag{4a}$$

$$\beta_{o\infty}^o = 1.1 \times 10^{-7} T^{-1/2} e^{-15.8/T}. \tag{4b}$$

For argon ions, $V_\infty = 27.5$ volts, we have

$$\beta_{ij}^+ = 1.8 \times 10^{-10} T^{-1/2} e^{-V_x/T}$$

(4c)

$$\beta_{o\infty}^+ = 1.8 \times 10^{-7} T^{-1/2} e^{-27.5/T}.$$

(4d)

The average cross sections for excitation and ionization of an atom already in an excited state (for example, the resonance level) seem to be an order of magnitude greater than those of an atom in the ground state.[3] For the most part, this is due to the fact that $V_j - V_i$ is much smaller than $V_j - V_o$. Possibly there is an increase in σ_m because of the greater size of the excited atom, but for order-of-magnitude estimates we shall ignore this possibility.

3. Densities of Singly and Doubly Ionized Ions

It is necessary to know the densities of neutrals, singly and doubly ionized ions, and electrons in order to compute the rate of excitation from the ground state and rate of volume recombination. To estimate these densities, consider the following model. The atom has first and second ionization potentials V_1 and V_2, respectively. Neglect excitation and higher degrees of ionization. Suppose that ions are generated only by electron impact from the ground state, that is, $A \to A^+$ and $A^+ \to A^{++}$, with ionization coefficients β^o and β^+, respectively. Assume that recombination is at the walls, with a time constant τ for either ion.

In the steady state, the particle balance equations are

$$\frac{dn_o}{dt} = 0 = \frac{1}{\tau}(n_+ + n_{++}) - \beta^o n_- n_o$$

$$\frac{dn_+}{dt} = 0 = \beta^o n_- n_o - \left(\frac{1}{\tau} + \beta^+ n_-\right) n_+$$

(5)

$$\frac{dn_{++}}{dt} = 0 = \beta^+ n_- n_+ - \frac{1}{\tau} n_{++}.$$

The total heavy-particle concentration $N = n_o + n_+ + n_{++}$ is fixed; charge neutrality requires $n_- = n_+ + 2n_{++}$. Solving Eqs. 5, subject to these constraints, yields

$$\frac{n_-}{N} = \left[1 - \frac{k_1 + k_2}{2Nk_1 k_2}\right] + \sqrt{\left[1 - \frac{k_1 + k_2}{2Nk_1 k_2}\right]^2 + \frac{Nk_1 - 1}{N^2 k_1 k_2}}$$

$$\frac{n_o}{N} = \frac{1}{1 + k_1 n_-}$$

(6)

$$\frac{n_+}{N} = \frac{k_1 n_-}{(1 + k_2 n_-)(1 + k_1 n_-)}; \quad \frac{n_{++}}{N} = \frac{k_1 k_2 n_-^2}{(1 + k_2 n_-)(1 + k_1 n_-)},$$

where $k_1 = \beta^o \tau$, $k_2 = \beta^+ \tau$.

The excitation coefficients are given in Eq. 4b and 4d. The mean loss time for ions is difficult to evaluate. It must be somewhere between the free-flight loss time to the ends of the arc column and the radial diffusion time across the magnetic field. In the arc, this is between 10^{-5} and 10^{-2} sec. In Fig. XI-2, n_o, n_+, and n_{++} are plotted for $\tau = 10^{-3}$ and 10^{-4} sec, respectively, for $N = 10^{14}$ cm^3. If triply ionized atoms were included, the n_{++} density would reach a maximum and then fall off, and n_{+++} would build up as T_- increases. It is important to note that for T_- just roughly 0.5 volt above threshold, n_{++}/n_+ is 5-15 per cent, which value is an appreciable fraction of the total ionization.

The actual operating point of the arc can be varied by adjusting the arc current, gas-flow rate, and magnetic field. It should also be remembered that the plasma is also generated inside the hollow cathode where the density and temperature are much higher. It is, therefore, quite possible that the electron temperature in the positive column is somewhat lower than that calculated on this model for a given electron density.

4. Imprisonment of Resonance Radiation

If resonance radiation is strongly absorbed by atoms or ions in the ground state, it is imprisoned and leads to increased population of the resonance levels. Here we take the resonance levels to be the lower excited states, and neglect the higher states that have alternate transitions available. To estimate the degree of imprisonment, let the excitation of a resonance level be by electron impact and photon absorption, and let de-excitation by all processes except spontaneous emission have a time constant τ_d. Then, the balance equations for atoms in the resonance level and photons are

$$\frac{dn_1}{dt} = 0 = \beta_{o1} n_o n_- + B\rho n_o - \frac{n_1}{\tau_d} - An_1 \tag{7}$$

$$\frac{dn_\nu}{dt} = 0 = An_1 - B\rho n_o - \frac{n_\nu}{\tau_\nu}, \tag{8}$$

where A and B are the Einstein coefficients, ρ is the monochromatic energy density for resonance radiation, n_1 is the resonance level population, n_ν is the total photon density in the line, and τ_ν is the photon escape time. We shall approximate ρ by $n_\nu h\nu/\Delta\nu_D$, where $\Delta\nu_D = \frac{\nu}{c} \sqrt{\frac{2kT_o}{M}}$ is the Doppler width of the line.

Holstein[4] has given an expression for τ_ν in the limit of large absorption for cylindrical geometry and Doppler-broadened radiation:
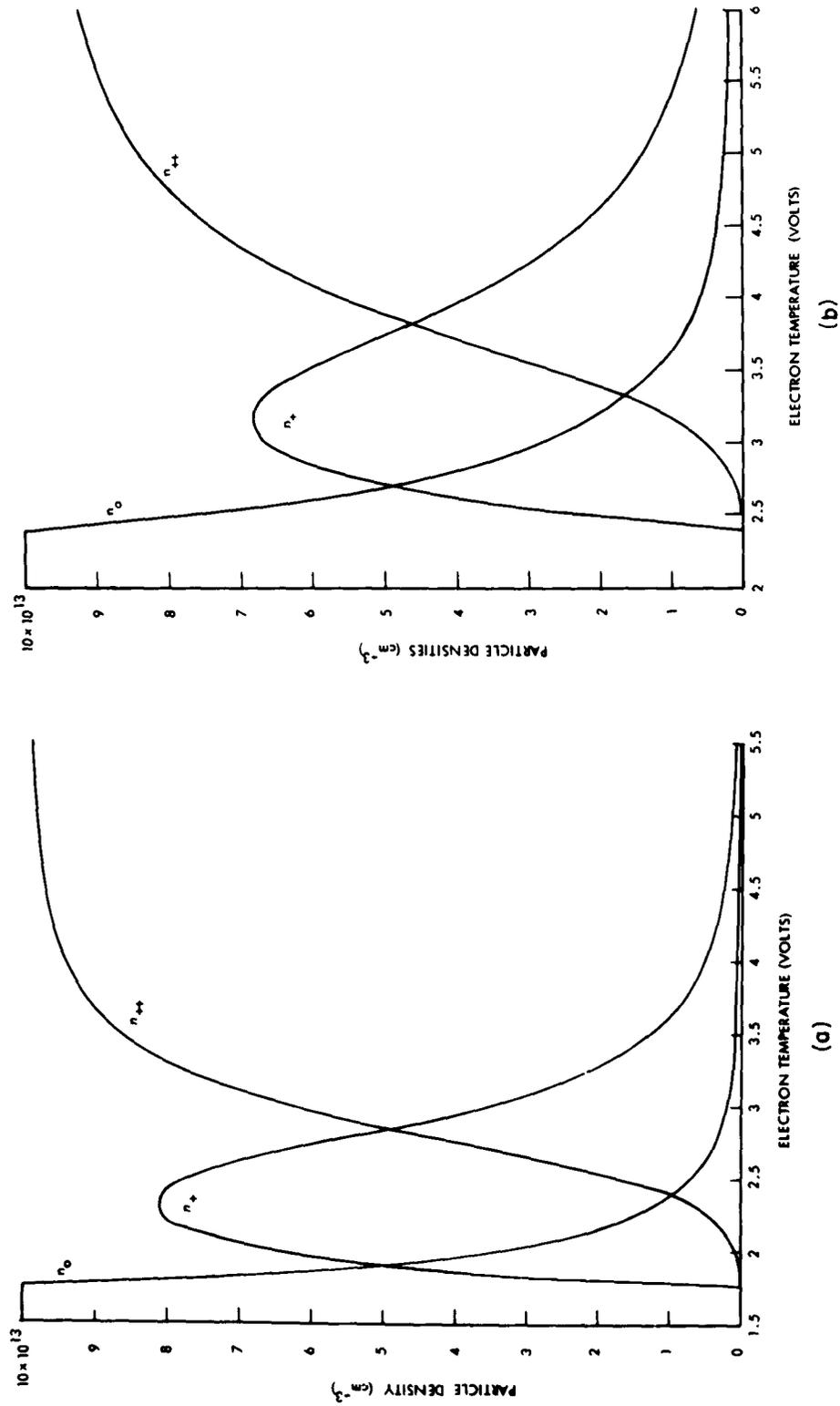
Fig. XI-2. Particle densities in highly ionized plasmas. (a) Ion loss time $\tau = 10^{-4}$ sec.
(b) Ion loss time $\tau = 10^{-3}$ sec.

$$\tau_\nu = \frac{k_o R \sqrt{\pi} \ln k_o R}{1.6A}, \tag{9}$$

where

$$k_o = \frac{1}{8\pi^{3/2}} \frac{c^2}{\nu^2} \frac{A g_1}{g_o} \frac{n_o}{\Delta\nu_D} \tag{10}$$

is the absorption coefficient at the center of the line.

If we introduce into Eq. 7 an imprisonment factor defined by the ratio of absorption to emission, $\mathcal{J} = B\rho n_o/A n_1$, we can see that the density n_1 is governed by spontaneous emission, or other processes accordingly as $(1-\mathcal{J})A\tau_d$ is much greater or much less than one. From Eqs. 8-10 and the ratio of Einstein coefficients, we obtain

$$\mathcal{J} = \frac{1}{1 + c \dfrac{T_o}{n_o^2}}, \tag{11}$$

where $c = 2.3 \times 10^{24}$ cm^{-6} volt^{-1} for both neutrals and ions, if we take $A = 5 \times 10^7$ sec^{-1}, $g_o/g_1 = 1/4$, $\lambda = 1050$ Å for the atom, and $A = 6.6 \times 10^8$ sec^{-1}, $g_o/g_1 = 1$, $\lambda = 700$ Å for the ion (rather arbitrarily), and $R = 3$ cm, $\sqrt{\ln k_o R} = 2$. The transition probabilities are uncertain, but are probably of the correct order of magnitude.

The value of $1 - \mathcal{J}$ for ions in the arc is roughly bracketed by 9×10^{-5} ($T_+ = 2^V$, $n_+ = 10^{13}$ cm) and 4.5×10^{-2} ($T_+ = 0.1^V$, $n_+ = 5 \times 10^{13}$ cm^3). The other principal depopulating processes are likely to be either superelastic or ionizing collisions with electrons. For these, τ_d may be between 10^{-4} and 10^{-5} sec. Thus we might expect $(1-\mathcal{J})A\tau_d$ to be between 0.6 and 3×10^3. This means that the imprisonment is generally not quite great enough for us to neglect spontaneous emission as a depopulating process as compared with the other losses, but it is enough to decrease the effective transition probability by several orders of magnitude.

5. Competing Processes

We now consider the relative rates of excitation and de-excitation of a nonresonant level. Neglect photon absorption and populating collisions of the second kind. Assume that stepwise excitation is from the resonant level only, and that cascading is from a single upper level. The balance equation is

$$\frac{dn_j}{dt} = 0 = \beta_{oj} n_- n_o + \beta_{1j} n_- n_1 + A_{kj} n_k + \epsilon_j n_\infty n_- - n_j [A_j + \tau_{d_j} + \tau_{c_j} + \beta_{jk} n_- + \beta_{j\infty} n_-], \tag{12}$$

where n_o, n_1, n_j, n_k are the populations of the ground, resonant, radiating, and cascading levels, respectively; n_∞ is the ground-state density of ions of next higher degree; α_j is the recombination coefficient; and τ_{d_j} and τ_{c_j} are the mean times for escape from the plasma and collisions of the second kind, respectively. Typical values of all coefficients are given in Table XI-1 for the argon ion, with $V_1 = 16^V$, $V_j = 20^V$, $V_k = 22^V$, and $V_\infty = 27.5^V$.

It is quite apparent that the upper levels de-excite almost exclusively by spontaneous emission, as expected.

Radiative recombination is the only significant volume recombination process in the arc. The partial coefficient for recombination to an individual level decreases rapidly with distance above the ground state, especially when the electron energy is high.[5] The formula $\alpha_j = 10^{-13} T^{-3/4}$ represents only a crude estimate based on calculations for monatomic hydrogen.[3,5] In Table XI-2 we give the ratio of recombination rate to electron excitation rate, using the formulas in Table XI-1, and taking n_{++}/n_+ from Fig. XI-2a. It is seen to be a marginal proposition to neglect this process, especially at high arc currents, when n_+ begins to decrease.

We can obtain an upper limit to the ratio of electron excitation from the resonance level to that from the ground state by assuming 100 per cent imprisonment and depopulation of the resonance level by ionization by electron impact. This ratio is

$$\frac{\beta_{1j} n_1}{\beta_{0j} n_0} \approx \frac{\beta_{1j} \beta_{01}}{\beta_{0j} \beta_{1\infty}} = 10^{-3} e^{(V_\infty - V_1)/T_-}$$

(see Table XI-2). Incomplete imprisonment and other resonance level losses — particularly superelastic collisions — tend to decrease this ratio.

If we assume that the cascade level is populated only by electron excitation from the ground state and depopulated only by spontaneous emission, the ratio of cascade to excitation contributions to n_j is simply

$$\frac{A_{kj} n_k}{\beta_{0j} n_- n_0} \approx \frac{A_{kj}}{A_k} e^{-(V_k - V_j)/T_-}.$$

Now, A_{kj} is proportional to $(V_k - V_j)^3$, so we would expect the branching ratio A_{kj}/A_k to be close to one only for transitions at optical frequencies. Thus we let $V_k - V_j \approx 2^V$ in the comparison in Table XI-2. Again, we have a marginal situation, but the estimates are presumably on the high side.

6. Intensity Measurements

A Jarrell-Ash 0.5-meter scanning monochromator was used in measuring the intensities of the argon ion lines. In the initial survey runs, light from the entire arc was

Table XI-1. Values of population coefficients.

Coefficient	$T_- = 2^V$	4^V	6^V
$\beta_{oj} = 1.8 \times 10^{-10} T_-^{-1/2} e^{-20/T_-}$ cm³/sec	5.8×10^{-15}	6.1×10^{-13}	2.6×10^{-12}
$\beta_{ij} = 1.8 \times 10^{-10} T^{-1/2} e^{-4/T_-}$	7.1×10^{-12}	3.31×10^{-11}	3.8×10^{-11}
$\beta_{jk} = 1.8 \times 10^{-10} T^{-1/2} e^{-3/T_-}$	2.8×10^{-11}	4.3×10^{-11}	4.5×10^{-11}
$\beta_{j\infty} = 1.8 \times 10^{-7} T^{-1/2} e^{-7.5/T_-}$	3.0×10^{-9}	1.4×10^{-8}	2.1×10^{-8}
$a_j = 10^{-13} T_-^{-3/4}$ cm³/sec	6.0×10^{-14}	3.5×10^{-14}	2.6×10^{-14}
$A_{kj}, A_j \sim 10^9$ sec⁻¹			
$\frac{1}{\tau_{dj}} \sim 10^2 - 10^4$ sec⁻¹			
$\frac{1}{\tau_{cj}} \sim 10^4$ sec⁻¹			

Table XI-2. Ratios of secondary processes to electronic excitation.

Ratio	$T_- = 2$	2.4	3
$\dfrac{a_j n_\infty}{\beta_{oj} n_o} = 5.5 \times 10^{-4} T^{-1/4} e^{20/T_-} \dfrac{n_{++}}{n_+}$	.097	.24	.60
$\dfrac{\beta_{ij} n_1}{\beta_{oj} n_o} = 10^{-3} e^{11.5/T_-}$	.32	.12	.046
$\dfrac{A_{kj} n_k}{\beta_{oj} n_o n_-} = e^{-2/T_-}$	.37	.48	.51

allowed to fall on the entrance slit, which was perpendicular to, and approximately 2 feet from, the arc axis. Later, a system of collimating slits and mirrors was interposed between the arc and the monochromator. This arrangement, Fig. XI-3, takes a
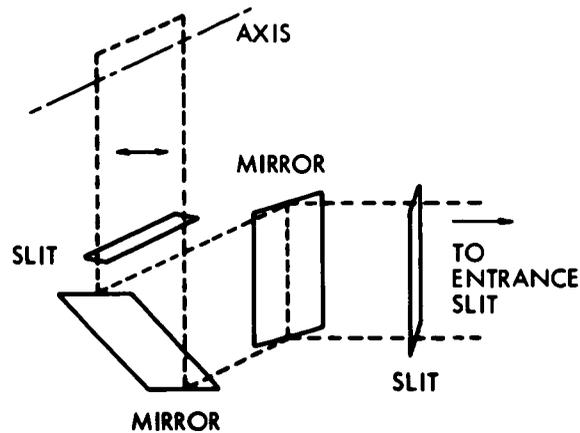
Fig. XI-3. Movable collimator.

narrow ribbon of light (0.005 in.) parallel to the arc axis and rotates it 90° so that it falls on the monochromator entrance slit. A lens was used to focus the light on the slit. The collimating-slit system is movable, scanning the arc perpendicular to the axis. No provision has been made for absolute intensity measurements; all measurements are in arbitrary units. A tungsten-ribbon filament lamp is being added to a new arrangement for calibration of intensities. The 1P21 photomultiplier output was amplified by a Philbrick UPA-2 operational amplifier, and displayed on a Varian G11A strip chart recorder. Often the intensity variation in one run would be large enough to necessitate changing the amplifier gain or photomultiplier voltage to keep the recorder on scale. In these cases, the run was made in overlapping segments, and the raw data were plotted on log paper. The gain factor was obtained from the displacement of the curves.

The measurements performed thus far have been on

(a) **Light from the Whole Arc**

  (i) A spectrogram from 3200 Å to 5200 Å.

  (ii) Intensity versus arc current for several ion lines.

  (iii) Intensity versus magnetic field for several ion lines.

(b) **Ribbon of Light (Arc Profiles)**

  (i) Arc profiles of several ion lines and many arc conditions.

(ii) Intensity on axis versus current.

A discussion of these measurements follows.

a. i. Approximately 30 AI, 190 AII, and 7 AIII lines have been identified out of more than 300 argon lines in a spectrogram taken at 20 amps and 800 gauss. All of the AI and AIII lines are rather weak compared with the most intense lines of the single ion. Twelve, out of 40, unidentified impurity lines had measurable, but very small, intensities. They are possibly Ta, Ne, or Fe lines, but the wavelengths are not precise enough for us to be sure. The intensity of the strongest impurity line ($\lambda 4580$; Ta 4580.7 ?) is approximately 10^{-3} times the intensity of the stronges AII line (λ 4348).

Ion lines, strong enough to be of use, are listed below in order of increasing excitation potential.

V_x	λ	I (arb)	V_x	λ	I (arb)
19.14	4806	25,000	19.88	3851	9,550
19.18	4736	22,800	21.05	4609	16,600
19.41	4348	25,000	21.40	4370	11,800
19.46	4426	25,000	22.85	3588	5,400
19.88	3729	9,800	23.07	3868	1,670

a. ii. The intensities of $\lambda 4806$, $\lambda 4736$, $\lambda 4348$, $\lambda 4426$ are plotted in Fig. XI-4 as I_1, those of $\lambda 4370$, $\lambda 3588$ as I_2, all normalized to 100 at 65 amps. The magnetic field was 1200 gauss. The flow rate was 130 cc-atm/min which produced a pressure of 2-3 μ Hg in the arc tube. Although there is considerable scatter in the data points, there is a distinct difference between the intensities of lines originating at levels between 19.1 volts and 19.5 volts and those originating between 21.4 volts and 22.9 volts.

If we assume that the line intensities are given by

$$I_j = c_j n_- n_+ T^{-1/2} e^{-V_j/T} \qquad j = 1, 2,$$

and take logarithms of intensity ratios, we can see how T_- varies once it is known for one current setting. $V_2 - V_1$ is approximately equal to 3 volts. Taking ratios at 30, 65, and 100 amps, and assigning values to T_{100}, we obtain the values listed below

T_{30}	T_{65}	T_{100}
1.9	2.2	2.4
2.1	2.4	2.6
2.2	2.5	2.8
2.3	2.8	3.0
2.5	3.2	3.5
2.9	3.6	4.0
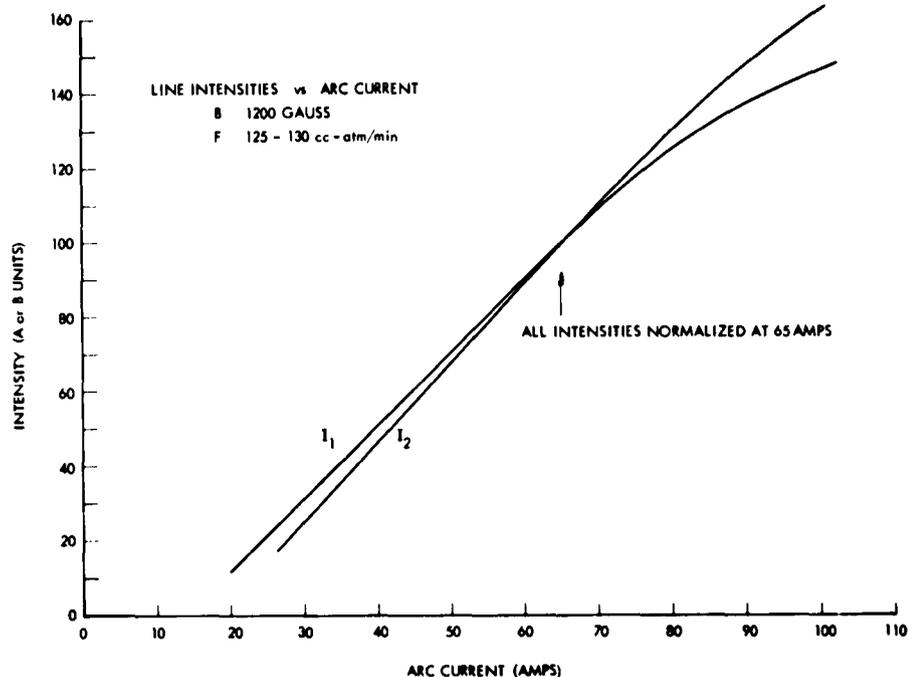
Fig. XI-4. Line intensities vs arc current.

The fact that the intensity of ion lines increases linearly with current is rather surprising, since it is proportional to electron density times ion ground-state density. Since, even at 30 amps, the concentration of n_{++} is great enough to produce significant AIII line intensity, it is probable that the arc is operating near the peak of the n_+ density. The slower increase in intensity at currents higher than 80 amps could then be explained by a depletion in n_+ density as n_{++} builds up; comparison of this table with Fig. XI-2 supports this hypothesis.

a. iii. All line intensities increased gradually as the magnetic field was increased from 1000 gauss to 2200 gauss, at 65 amps arc current. Total increase was appoximately a factor of 2, and no significant dependence on excitation potential could be discerned.

b. i. A typical profile is shown in Fig. XI-5. This curve is λ4348 at 50 amps, 1200 gauss, and 100 cc-atm/min (1-1.5 μ Hg in arc tube). The original idea in this experiment was to measure the radial variation of intensity for three lines with widely separated upper levels, for example, λ4348, λ4370, λ3868 at 19.4, 21.4, 23.1 volts, and deduce the radial variation in electron temperature, by comparing intensity ratios as we did in (a. ii). This experiment has not yielded useful results because accurate alignment of the monochromator could not be maintained because
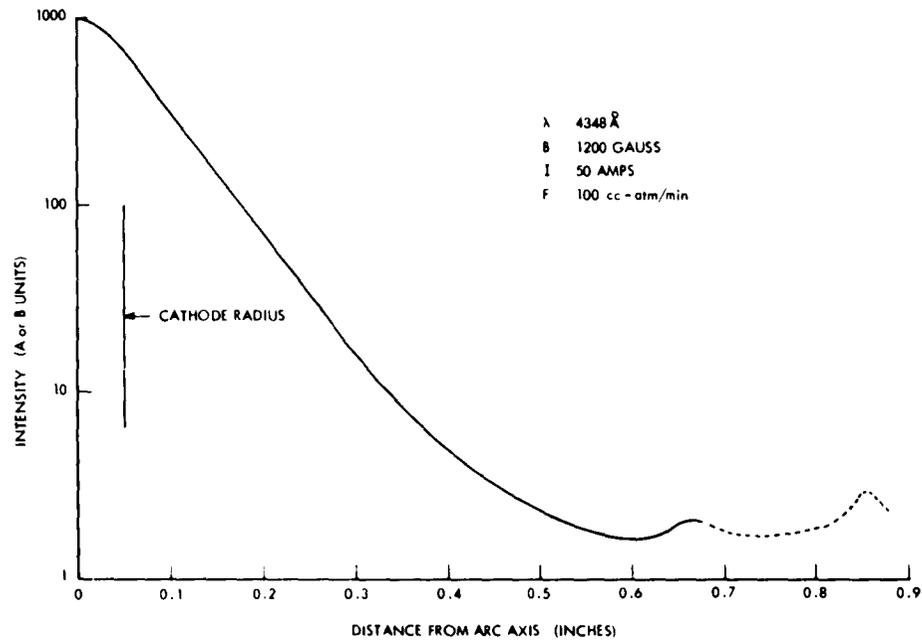
Fig. XI-5. Intensity of λ4348 vs distance from arc axis.

of vibration; also, the arc does not remain sufficiently constant from run to run.

A new monochromator and slit-system mount has been made to solve the alignment problem, and we shall try again. A moving probe is now being built which should sweep through the arc fast enough to enable density and other probe measurements without burning up. Once we know, from probe measurements, the plasma density profile, and the electron temperature close to the edge of the arc, we shall be able to use a single line intensity curve to obtain the temperature profile.

Two features of Fig. XI-3 are worth noting:

(i) The intensity falls off almost exponentially with distance from arc axis — not radius, as the data have not yet been inverted. The slope of the log I curve then affords a measure of the width of the arc column, as determined by diffusion. The variation of width (reciprocal of the slope) with magnetic field is shown for several conditions in Fig. XI-6. If anything, there seems to be an increase of a few per cent in width as B runs from 800 gauss to 2000 gauss. There is a much more pronounced increase as current and/or flow rate are increased, but no systematic measurements have been made.

(ii) The width of the crown — the portion from the axis to the exponential — is not determined solely by the cathode diameter. At higher pressure it broadens
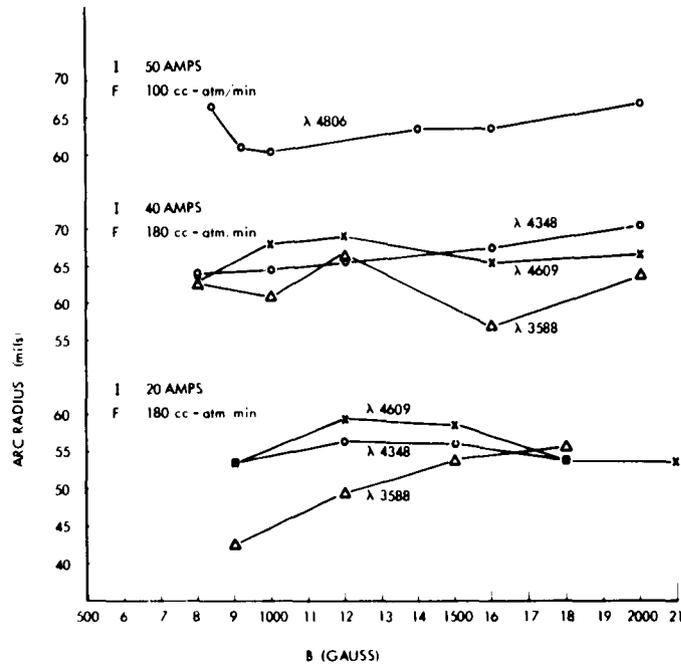
Fig. XI-6. Arc width vs magnetic field.

considerably and the exponentials are flatter. Also, we looked for a bump in this crown which might be due to excitation by primary electrons from the cathode. None was found within the resolution of the measurements, which is perhaps a few per cent near the intensity maxima. These observations lend further support to our general analysis, which assumes that all of the light is generated by plasma electrons rather than by a monoenergetic beam issuing from the cathode.

b. ii. Intensity versus current measurements have to be repeated with the collimating slit set on-axis because the alignment problem became worse when these measurements were made. Generally speaking, however, the intensity seems, first, to rise and then to drop off rather sharply as the current is increased. This indicates that the n_+ is markedly depleted, and n_{++} is considerably increased near the axis at high current.

C. D. Buntschuh

References

1. C. D. Buntschuh, Quarterly Progress Report No. 65, Research Laboratory of Electronics, M.I.T., April 15, 1962, pp. 79-87.

2. L. M. Lidsky, S. D. Rothleder, D. J. Rose, S. Yoshikawa, C. Michelson, and J. R. Mackin, Jr., J. Appl. Phys. 33, 2490 (1962).

3. H. S. W. Massey and E. H. S. Burhop, Electronic and Ionic Impact Phenomena (Oxford University Press, London, 1952).

4. T. Holstein, Phys. Rev. 72, 1212 (1947); 83, 1159 (1950).

5. R. Fowler, Radiation from Low-Pressure Discharges, Handbuch der Physik, Vol. 22, Edited by S. Flügge (Springer Verlag, Berlin, 1956), pp. 209-253.

B. INDIUM ANTIMONIDE PHOTODETECTOR FOR SUBMILLIMETRIC RADIATION

This photoconductive detector, purchased from Mullard Ltd., is sensitive to radiation from 100 μ to beyond 1 mm, and has a response time of less than 1 μsec. The InSb sample, located at the bottom of a vertical light pipe, is immersed in liquid helium. By pumping on the helium, the sample temperature is reduced to less than 1.6°K. The InSb operates in a 6 kgauss magnetic field obtained from a superconducting niobium solenoid.

The InSb detector was compared with a Golay cell. Both detectors were exposed to radiation from 200 μ to 300 μ, chopped at 15 cps. The signals from each detector were preamplified, and then synchronously detected. The signal-to-noise ratio of the InSb detector was approximately 25 times better than that of the Golay cell. Further improvement in signal-to-noise ratio has resulted from increasing the chopping frequency to 3100 cps, and from a corresponding decrease in preamplifier noise.
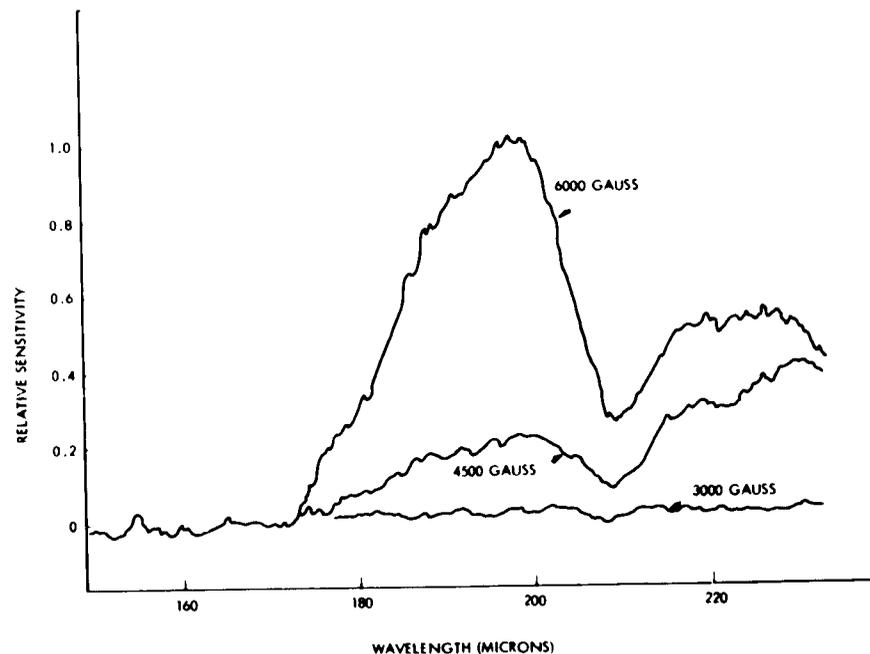
Fig. XI-7. InSb detector sensitivity as a function of wavelength magnetic field surrounding InSb sample varied from 3 kgauss to 6 kgauss.

The detector exhibited a short wavelength cutoff, as shown in Fig. XI-7, which is dependent on the magnetic field strength.

R. E. Whitney

C. PLASMA-WAVE COUPLING ANGLE

The dispersion relation for longitudinal plasma waves is

$$\left(\frac{A}{B-n^2}\right)_+ + \left(\frac{A}{B-n^2}\right)_- = 1. \tag{1}$$

Here, the plus and minus signs refer to ions and electrons.[1] The following definitions have been used: n is the refractive index, $A = \frac{\alpha^2}{\delta}$, $B = \frac{1-\beta^2}{(1-\beta^2 \cos^2 \theta)\,\delta}$, $\delta = \frac{kT}{mc^2}$, $\alpha^2 = \frac{\omega_p^2}{\omega^2}$, $\beta = \frac{\omega_b}{\omega}$, ω_p is the plasma frequency, ω_b is the cyclotron frequency, and θ is the angle between the propagation vector and the constant magnetic field. The ion flux and electron particle flux along the k vector, $\Gamma_\pm$, for the two modes of plasma waves are related by the expression,

$$\left(\frac{\Gamma_-}{\Gamma_+}\right)_1 \cdot \left(\frac{\Gamma_-}{\Gamma_+}\right)_2 = -\frac{T_+}{T_-}, \tag{2}$$

in which the numeral subscripts refer to the modes. Equation 2 states that if the electrons and ions vibrate in phase for one mode, they will be out of phase for the other. The angles for which the phase of vibration changes are the resonance angles given by

$$\tan^2 \theta_R^+ = \beta_+^2 - 1, \qquad \tan^2 \theta_R^- = \beta_-^2 - 1. \tag{3}$$

These results have been derived and discussed elsewhere.[1,2]

Equation 1 can be written in a form that displays the coupling of ion and electron waves:

$$\left(n^2 + A_+ - B_+\right)\left(n^2 + A_- - B_-\right) = A_+ A_-. \tag{4}$$

If the coupling term, $A_+ A_-$, is disregarded, an angle is determined at which the separated ion and electron waves intersect on a polar plot of phase velocity as a function of angle of propagation. One would suspect that this angle would indicate a region of strong coupling between the waves and therefore have physical significance. This coupling angle, θ_c, is defined as

$$A_+ - B_+ = A_- - B_-. \tag{5}$$

When ion and electron temperatures are equal, Eq. 5 simplifies considerably and yields

$$\tan^2 \theta_c = \frac{\left(\beta_-^2 - 1\right)\left(1 - \beta_+^2\right)}{1 + \beta_+ \beta_-}.$$ (6)

Hence such a coupling angle will exist only between the cyclotron lines of the α^2, β^2 plane.[2]

If we consider the kinetic energy density (W) caused by motion along the propagation direction, that is,

$$W = \frac{m\Gamma^2}{2N_o} = \frac{1}{\alpha^2}\left(\frac{A}{B - n^2}\right)^2 \frac{E_k^2}{8\pi},$$ (7)

where $\Gamma^2 = \frac{\Gamma\Gamma^*}{2}$, and $\frac{E_k^2}{8\pi}$ is the energy density of the longitudinal electric field, and make use of Eq. 2, we obtain

$$\left(\frac{W_-}{W_+}\right)_1 \left(\frac{W_-}{W_+}\right)_2 = \left(\frac{\delta_+}{\delta_-}\right)^2.$$ (8)

In the very special case $\delta_+ = \delta_-$, that is, for equal electron and ion thermal velocities, Eq. 8 states that when the ions dominate the longitudinal kinetic energy of one mode, the electrons will dominate the other. For this case, the angle at which the kinetic energies of a particular mode change from electron- to ion-dominated will be the coupling angle θ_c. In the general case, however, there is no sharp distinction in the kinetic energy of the modes.

We shall now examine the potential energy density (which has been given by Allis, Buchsbaum, and Bers[3]):

$$\phi = \frac{1}{4} \frac{mV_T^2}{N_o} N_1 N_1^* = \frac{1}{4} \frac{mV_T^2}{N_o} \left(\frac{k}{\omega}\right)^2 \Gamma_k \Gamma_k^*,$$ (9)

where V_T is the thermal velocity; ϕ is $-\frac{P \cdot \delta V^*}{2V}$, with $P = N_1 kT = 1/2 N_2 m V_T^2$, the pressure, and $\frac{\delta V}{V} = -\frac{N_2}{N_o}$ the fractional volume change.

Again using Eq. 2, we find the following relation for potential energies:

$$\left(\frac{\phi_-}{\phi_+}\right)_1 \left(\frac{\phi_-}{\phi_+}\right)_2 = 1.$$ (10)

Equation 10 indicates that if the potential energy is ion-dominated for one mode, it will be electron-dominated for the second mode. Furthermore, if we set $\phi_+ = \phi_-$, we find from Eqs. 7 and 9 that

$$n^4 + 2n^2\left[\frac{A_+B_- - A_-B_+}{A_- - A_+}\right] + \frac{A_-B_+^2 - A_+B_-^2}{A_- - A_+} = 0.$$

(11)

Equation 4 when written as a quadratic in n^2 is

$$n^4 + n^2[A_-+A_+-B_--B_+] + B_-B_+ - B_+A_- - B_-A_+ = 0.$$

(12)

Equating either the constant or middle coefficients of Eqs. 11 and 12 results in Eq. 5. Therefore, the potential-energy interchange occurs at the coupling angle. Thus a physical interpretation has been found for the geometrical coupling between the independent electron and ion plasma waves.

H. R. Radoski

References

1. H. R. Radoski, Dispersion relation for longitudinal plasma waves, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, p. 100.

2. W. P. Allis, S. J. Buchsbaum, and A. Bers, Waves in Anisotropic Plasmas (The M.I.T. Press, Cambridge, Mass., 1963), Chapter 5.

3. Ibid., Chapter 8, p. 119.

A. BEAM-PLASMA DISCHARGES

1. SYSTEM A: MAGNETIC PROBE

Several magnetic probes have been constructed and used to aid in the diagnosis of the beam-plasma discharge (Fig. XII-1). The coils are 100 turns of No. 36 wire with an
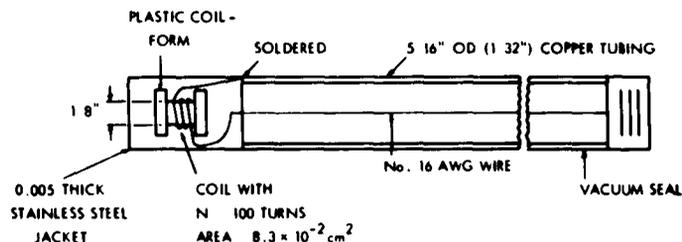
Fig. XII-1. Magnetic probe.

area of 8.3×10^{-2} cm^2. The coils are constructed with their axis either parallel or perpendicular to the axis of the coaxial line. The relationship between the voltage at the terminals and the changing magnetic field through the coil is expressed in Eq. 1.

$$V = NAg \frac{dB}{dt},$$
(1)

where g is a frequency-dependent factor that accounts for the skin effect of the stainless-steel jacket. This factor is found experimentally by calibrating the probe in the field of a solenoid driven by a sinusoidal current source (Fig. XII-2). The effect of the jacket
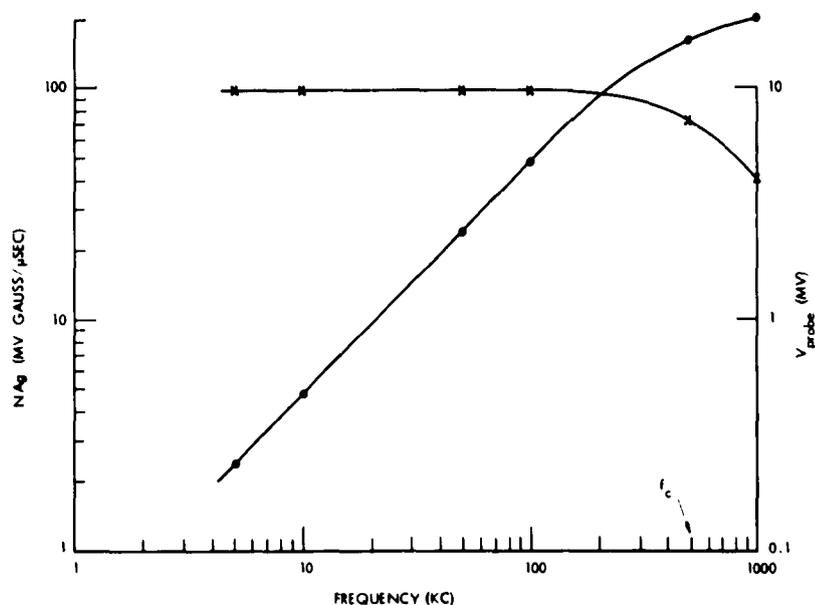
Fig. XII-2. Plot of V_{probe} vs frequency (...) and plot of NAg vs frequency (xxx) at constant current, 0.01 amp (1 amp = 7.8 gauss).

is to limit the high-frequency response of the coil. If the term dB/dt in Eq. 1 can be expressed as a series of impulse functions, the voltage response will be a series of scanning functions. To a good approximation, the width of the major lobe of the scanning function is equal to the inverse of the cutoff frequency defined in Fig. XII-2. If the impulse functions are far enough apart, the major lobes of the scanning functions will not overlap, and the area of the triangle circumscribed by the major lobe of the scanning function is equal to the area of the corresponding impulse.

Thus, in Fig. XII-3, we can relate the pulses in the magnetic probe voltage to step

MAGNETIC PROBE VOLTAGE

COLLECTOR CURRENT

TIME 20 μSEC/CM
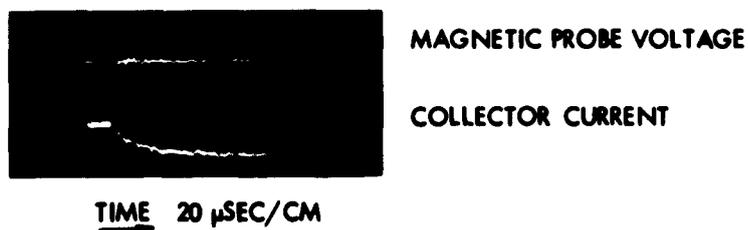
Fig. XII-3. Oscillogram of magnetic probe voltage and collector current. V_{beam} = 5 kv; pressure = 3×10^{-4} mm Hg hydrogen; B_0 = 320 gauss. Calibration: I_{coll} = 0.2 amp/cm; V_{probe} = 5 mv/cm.

changes in the local axial field of the plasma (the coil axis was aligned parallel with the dc magnetic field). Typically, the changes in the local field are approximately 0.5 gauss. The first pulse occurs simultaneously with the break in beam current, and the second pulse, usually smaller but longer in duration, occurs at the end of the beam current (Fig. XII-3).

Figure XII-4b shows that by turning the probe 90° with respect to the dc magnetic field, the pulses disappear. At low pressures (5×10^{-6} mm Hg) when the electron beam does not initiate a beam-plasma discharge, there is no voltage induced on the magnetic

a

b

V_{probe}

LIGHT

I_c

c
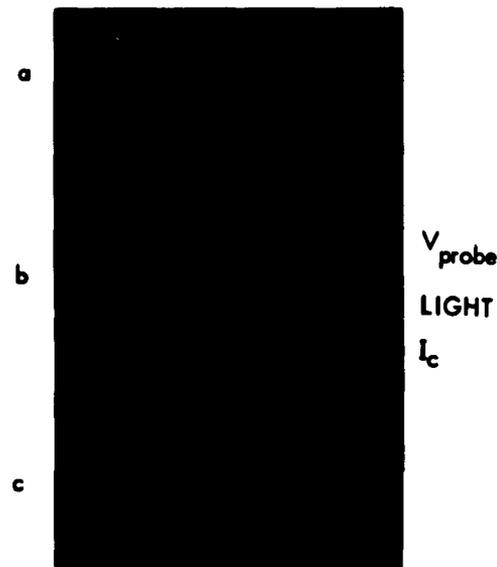
Fig. XII-4. Pulse structure and oscillatory nature of magnetic probe voltage. V_{beam} = 10 kv; pressure, 2.2×10^{-4} mm Hg hydrogen; B_o = 240 gauss.

(a) The magnetic probe is aligned with its coil axis parallel to the dc magnetic field. Calibration: 20 μsec/cm and 10 mv/cm.
(b) The magnetic probe is aligned with its coil axis radially perpendicular to the dc magnetic field. Calibration: 20 μsec/cm; 20 mv/cm.
(c) The magnetic probe voltage is expanded to 1 μsec/cm; 20 mv/cm.
In all three oscillograms, the light calibration is 50 μsec/cm; 5 volts/cm.
Collector current calibration: 50 μsec/cm; 0.5 amp/cm.

probe for any orientation of the probe axis. In Fig. XII-4a, the change in B_z measured from the pulse at the end of the beam-plasma discharge is 0.9 gauss.

Using the relationship

$$\frac{B\Delta B}{\mu_0} = e\Delta NT, \tag{2}$$

we obtain

$$\Delta NT = 5 \times 10^{24} \, B\Delta B. \tag{3}$$

If we assume a density of 10^{12}/cc, we can estimate an electron temperature of 11 ev.

In some experiments, it was observed that another set of pulses appears before the beam current ends (Fig. XII-5). This set of pulses can (tentatively) be interpreted as
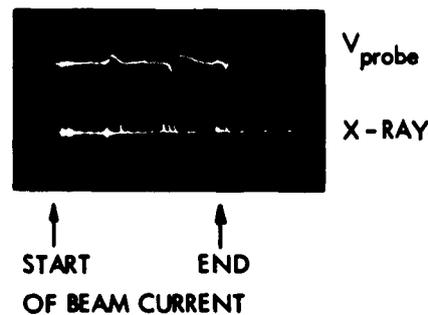
Figure XII-5.

"Collapse" and rebuild-up of beam-generated plasma. V_{beam} = 5 kv; pressure = 1.8×10^{-4} mm Hg hydrogen; B_0 = 350 gauss. Calibration: time, 20 μsec/cm; V_{probe}, 5 mv/cm; x-ray, 50 mv/cm.

the "collapse" and rebuilding of the beam-generated plasma. On a simultaneous display of x-rays in Fig. XII-5, it is seen that the x-rays precede such a "collapse" of the
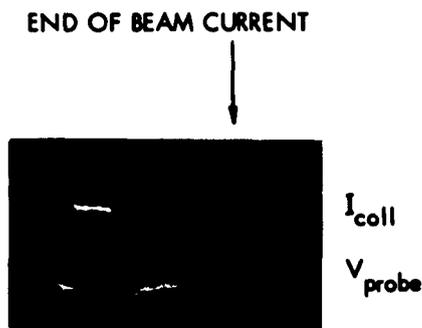
Figure XII-6.

Presence of 600-kc oscillation on collector current. V_{beam} = 10 kv; pressure = 2.2×10^{-4} mm Hg hydrogen; B_0 = 340 gauss. Calibration: time, 10 μsec/cm; V_{probe} = 5 mv/cm; I_{coll}, 20 ma/cm.

plasma. This phenomenon has not been studied in great detail.

A very clean sinusoidal oscillation sometimes with second harmonics, with a frequency of 450 kc to 1 mc has also been observed (Figs. XII-4 and XII-6). This oscillation occurs at high beam voltages (~10 kv). The frequency of the oscillation gradually decreases as the magnetic field is increased by a factor of 3. The frequency of the oscillation does not depend on the mass of the gas (argon and hydrogen were used) or on

pressure. At high pressures (~1μ) the amplitude of the oscillation decreases. The oscillations start after the first pulse appears, and may last long after the beam current has ceased (Fig. XII-6). The oscillations are also seen on other collectors (Fig. XII-6).

These oscillations have a standing-wave character. The half wavelength of a 450-kc oscillation found from a movable magnetic probe is 15 cm. This gives a phase velocity of 1.35×10^7 cm/sec, which is larger than the acoustic velocity (~10^6 cm/sec) and smaller than the Alfvén velocity (~10^8 cm/sec). The mechanism of the generation of these oscillations is being studied.

<div align="right">H. Y. Hsieh</div>

2. SYSTEM B: ROTATIONAL INSTABILITY AND QUENCHING

Studies of the beam-plasma discharge in the long solenoid apparatus (System B) have disclosed a low-frequency rotation of the plasma that is assumed to represent an instability or a mechanism by which the plasma leaks across the field. We have also observed

(a)

(b)
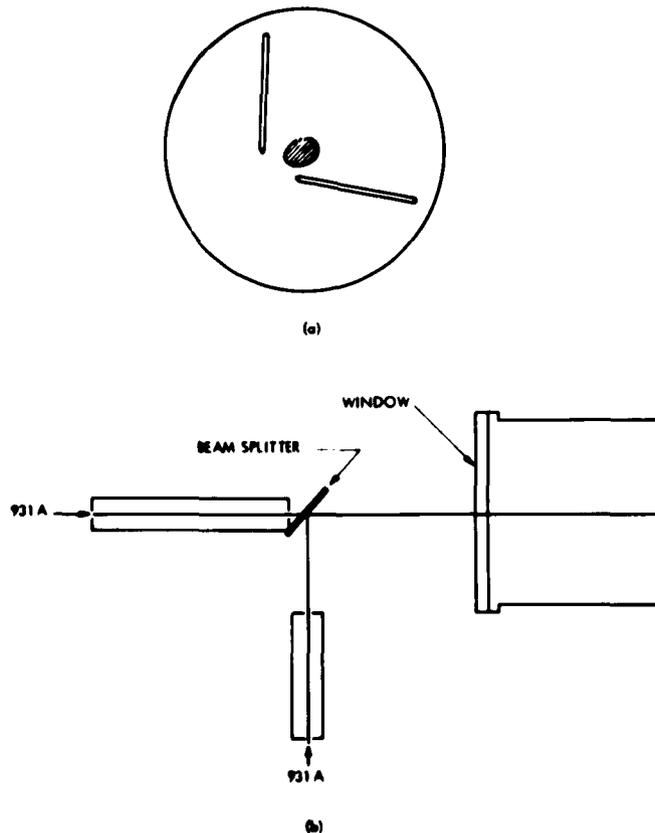
Fig. XII-7.   (a) Plasma current probes.
            (b) Space-time resolved light arrangement.

some critical values of magnetic field for which no beam-plasma interaction occurs.

a. Rotational Instability

Independent measurements of plasma currents and plasma light have been used to measure the rotation. The two experiments are illustrated in Fig. XII-7. Use of two small apertures for each photomultiplier and a beam splitter allowed simultaneous
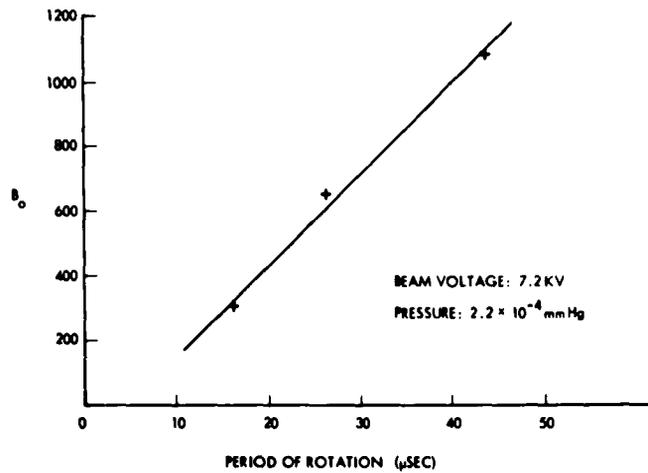
Fig. XII-8. Rotational period vs applied magnetic field.
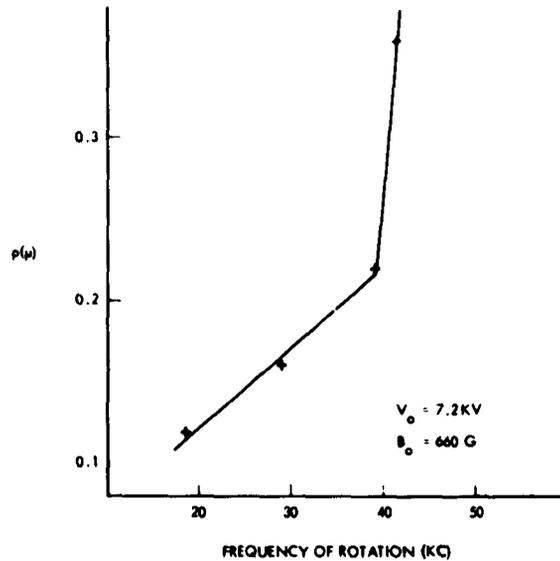
Fig. XII-9. Rotational frequency vs pressure.

measurements of light at difference positions in the cross section of the plasma column. The variation of rotational period with magnetic field (Fig. XII-8) suggests an E/B drift with E $\approx$ 2 volts/cm. The variation of rotational frequency with gas pressure is shown in Fig. XII-9. Work is in progress to determine if a helical mode exists and to determine dc space-charge fields.

b. Quenching of Beam-Plasma Discharge

Values of the applied magnetic field for which the beam-plasma discharge is extinguished have been measured. For critical values of the magnetic field the beam passes through the gas without exciting microwave oscillations, although low-frequency oscillations (a few megacycles) are present.[1] The "stability" depends on pulse length.
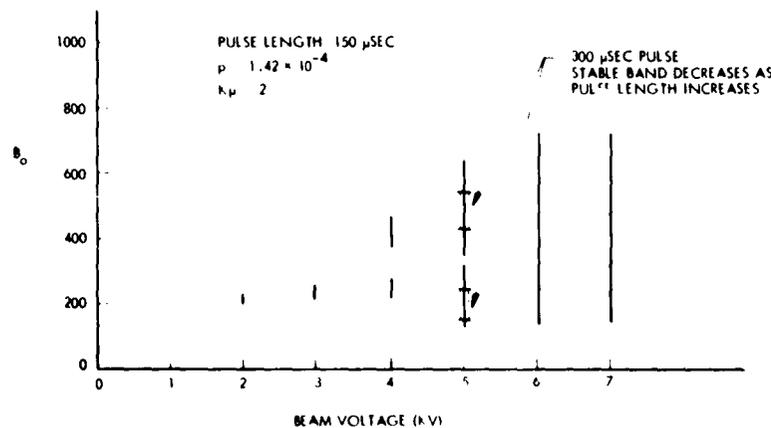
Fig. XII-10. Beam-gas stability. Solid line indicates band of stability.

Thus, a wide band of magnetic field strength produces stability for short pulses, but the range of B_o decreases as the pulse length increases. The stability depends on perveance and is more easily attained at lower pressures. The regions of stability at microperveance 2 are plotted in Fig. XII-10.

B. A. Hartenbaum

References

1. L. D. Smullin, W. D. Getty, B. A. Hartenbaum, and H. Y. Hsieh, Electron beam-plasma interaction experiments, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, pp. 116-123; see especially p. 123.

3. SYSTEM C: STRONG INTERACTION BETWEEN A HIGH-DENSITY, HOLLOW
   ELECTRON BEAM AND A PLASMA

For several years our experiments[1,2] on strong electron beam-plasma interactions
have indicated that high beam perveance is important. Evidence of this can be seen by
increasing the perveance by only 20-30 per cent. Such an increase usually results in
stronger x radiation, rf oscillations, and light output from the plasma. There are good
theoretical reasons for expecting this behavior. The one-dimensional theory[3] for
collision- or temperature-limited reactive-medium amplification predicts that the max-
imum amplification rate increases rapidly with the ratio n_b/n_p, where n_b and n_p are
the beam and plasma electron densities. Since the beam density is proportional to per-
veance for space-charge-limited flow, an increase in perveance would produce a greater
amplification rate and consequently a stronger interaction. More recent theoretical
work[4] has shown that a large value of n_b/n_p is important if low-frequency ion oscilla-
tions are to be excited. To investigate the interaction with a much denser beam than
used heretofore, we have constructed an experiment in which the interaction between
a high-perveance, hollow electron beam and a beam-generated plasma will be studied.
In this report we shall describe the apparatus and some observations made in our first
exploratory operation of the experiment.

The interaction takes place in a 6-inch diameter stainless-steel vacuum tube that is
coaxial with a solenoid. The solenoid has an inside diameter of 7 inches, and outside
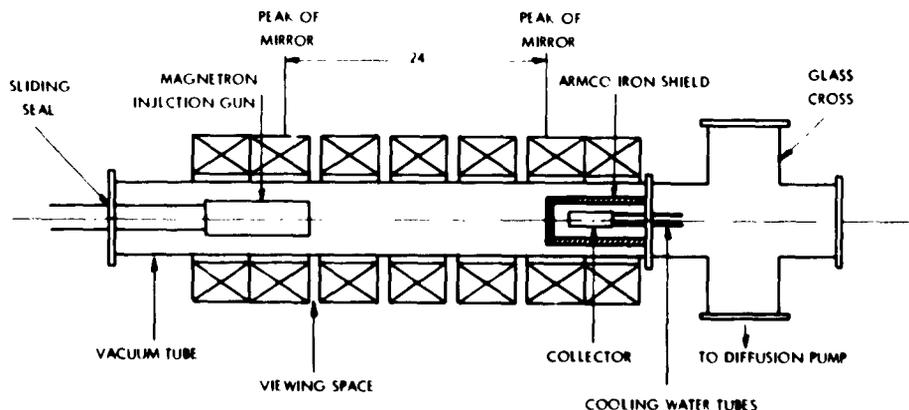
Fig. XII-11. Top view of hollow-beam experimental apparatus.

diameter of 13.5 inches, and is 40 inches long. The plasma can be observed
through 1-inch diameter quartz windows that can be viewed through spaces between the
coils of the solenoid. The vacuum chamber and solenoid are illustrated in Fig. XII-11.

The electron gun, which is also shown in Fig. XII-11, is mounted on a sliding tube that can be set at any desired axial position along the solenoid. All gun parts are non-magnetic. An Armco iron shield is mounted at the opposite end of the vacuum tube. This shield produces a magnetic mirror and magnetically shields the electron-beam collector. A second magnetic mirror is produced near the electron gun by running higher currents through two of the solenoid coils. We thus have a magnetic bottle with a 2:1 mirror ratio at each end. The distance between the peaks of the field is 24 inches.

To obtain a dense electron beam, we are using the magnetron injection gun designed and described by Poeltinger.[5] This gun is located at a peak of the mirror field. Its perveance ranges from 5×10^{-6} amp per $(volt)^{3/2}$ to 20×10^{-6} amp per $(volt)^{3/2}$, the range depending on the magnetic field and beam voltage. The outside diameter of the beam is approximately 0.5 inch, and the beam thickness is less than 10 per cent of its diameter. The electron density is 10^{11} electrons/cm^3.

An artificial-delay-line pulser has been built to drive the electron gun. This system is designed to operate up to 20 kv and 30 amps at a repetition rate of 10 pps. The pulse length is 300 μsec. A pulse transformer is used to match the 30-ohm line to the electron gun, whose nominal impedance is 500 ohms.

In the experiment the hollow beam is injected into the magnetic bottle near the peak of one of the mirrors. The beam diameter increases as it passes into the region of lower field between the mirrors. At the beam energies used in the experiment the beam passes through the second mirror and into the shielded collector, which is shown in Fig. XII-11. With the hydrogen pressure set in the range 10^{-4}-10^{-3} mm Hg in the vacuum tube, a plasma is rapidly generated by the beam. We have observed that this plasma strongly interacts with the hollow beam.

Our first exploratory experiments were made at beam voltages and currents up to 10 kv and 10 amps. The electron gun operated well in this range, and strong interaction was observed. Each beam pulse produced a very bright plasma whose color was light blue or light pink. At some operating conditions it was found that the light output was delayed 100-200 μsec after the start of the beam pulse. During this period, the current to the vacuum-tube walls was carried by ions and oscillated violently from zero to several amperes at frequencies of approximately 10 mc. Similar behavior has been observed previously.[2] A scintillator showed that x radiation escaped from the vacuum tube. Measurements with aluminum absorbers indicated that the x-ray energy, which was assumed to be monoenergetic, was in the range 17-20 kv. Observations of the microwave spectrum of the energy radiated by the plasma were made with the gun operating at 9 kv and 7.5 amps, and with the hydrogen pressure at 0.3 μ Hg. Strong microwave output was observed at frequencies as high as 32 kmc. The maximum oscillation frequency is assumed to equal the plasma frequency.[1] Strong oscillations were also observed at lower frequencies, particularly in the range 100-1500 mc. Plasma

oscillations at a plasma frequency of 32 kmc would correspond to a density of 10^{13} electrons/cm^3. The neutral gas density at the pressure of $0.3\,\mu$ Hg has approximately the same value, and therefore the plasma may be fully ionized.

W. D. Getty, L. D. Smullin

References

1. L. D. Smullin and W. D. Getty, Generation of a hot, dense plasma by a collective beam-plasma interaction, Phys. Rev. Letters 9, 3-6 (1962).

2. L. D. Smullin, B. A. Hartenbaum, and H. Y. Hsieh, Electron beam-plasma interaction experiments, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M.I.T., October 15, 1962, pp. 71-76.

3. G. D. Boyd, L. M. Field, and R. W. Gould, Excitation of plasma oscillations and growing plasma waves, Phys. Rev. 109, 1393 (1958).

4. R. J. Briggs and A. Bers, Electron-beam interactions with ions in a warm plasma, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M.I.T., April 15, 1963, pp. 65-74.

5. A. Poeltinger, High-perveance hollow electron-beam study, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M.I.T., July 15, 1962, pp. 25-29.

4. FAST-ACTING, CURRENT-ACTUATED GAS VALVE

The purpose of this investigation was to develop a valve that would admit repeatable quantities of gas (0.1-1 cc at S.T.P.) into a vacuum in pulses of less than 100 μsec. The valve that was developed is discussed in a Master's thesis[1]; therefore, only a brief discussion of its operation and characteristics will be given here.
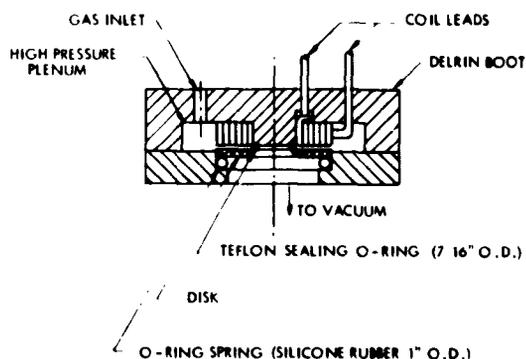
Fig. XII-12. Magnetically driven gas valve.

The valve shown in Fig. XII-12 consists of a 1-inch O.D. disk of copper, brass or aluminum which is held against the Teflon O-ring by the precompressed silicone rubber

O-ring. The 1-inch O.D. coil (1.7 μh) consists of nine turns of 0.25 × 0.021-inch copper tape, insulated with 0.004 Mylar ribbon and potted in epoxy. The 0.25-inch diameter coil leads are connected through a mechanical switch to a 2-μf capacitor that is initially charged to 3500-7000 volts. The ringing frequency of this circuit is approximately 90 kc, and its damping time constant varies between 4 μsec and 10 μsec, the variation depending on the material of the disk.

When the capacitor is discharged through the coil, the eddy currents induced in the disk cause a repulsive force between the coil and the disk. This force causes the disk to move downward and thus allows the gas to pass between the disk and Teflon O-ring, and then through the central hole in the disk. The energy-conversion efficiency of this device is approximately 0.5 per cent; however, the available force may be as high as 700-1000 pounds for a period of a few microseconds.

The period and amplitude of the disk motion may be accurately calculated by considering the valve as a damped spring-mass system in which the spring is precompressed and restrained from moving through its equilibrium position. The damping ratio is approximately $\rho = 0.88$ (in this case the damping is due to hysteresis in the O-ring spring). The period of the motion is not affected by variations in the initial capacitor voltage. However, when positive stops are added in order to limit the motion of the disk, the period varies not only with the stiffness and precompression of the O-ring but also with the initial capacitor voltage (the period decreases with increasing voltage) and the allowable excursion of the disk. In practice, the period could be varied 50-250 μsec, and the maximum amplitude was approximately 0.025 inch.

Finally, there are several difficulties that still must be corrected. The first is that the disk bounces on the Teflon O-ring, thereby reopening the valve. This bouncing is especially pronounced when high initial capacitor voltages are used with low precompression of the O-ring spring. This may be rectified by using a softer sealing O-ring. A second difficulty, excessive heating of the coil and disk, may be encountered if the valve is to be cycled rapidly for long periods of time.

D. S. Alles, L. D. Smullin

References

1. D. S. Alles, Design and Development of a Fast-Acting Valve for Plasma Research, S. M. Thesis, Department of Mechanical Engineering, M.I.T., May 1963.

B. INSTABILITY IN THE HOLLOW-CATHODE DISCHARGE

As previously reported,[1] a rotating instability has been observed in the hollow-cathode discharge. The instability causes a cloud of plasma, a "spoke," to rotate about the axis of the system, for sufficiently large values of axial magnetic field. This spoke

has fairly uniform properties in the direction of the magnetic field, and exhibits no heli-
cal pattern. During the past quarter, time-resolved, space-potential measurements
have been made, and a theoretical study has been initiated to describe macroscopic,
low-frequency instabilities of a magnetized plasma.

1. Effect of the Instability on Diffusion

Measurements of the average density at a point 4 inches from the system axis vs
magnetic field indicate that the presence of the instability is a major factor in the dif-
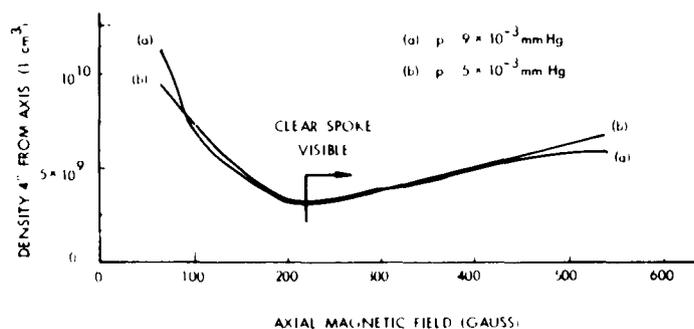
Fig. XII-13. Density 4 inches from system axis vs magnetic field. Clear spoke is
visible above ~220 gauss.
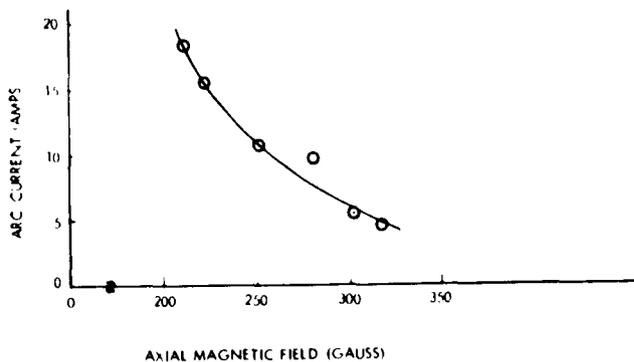
Fig. XII-14. Minimum axial magnetic field for instability vs
arc current.

fusion of plasma across the magnetic field (Fig. XII-13). The average density increases
with increasing magnetic field after the instability appears. Plasma density in the source
region remains nearly constant as the magnetic field is varied.

2. Onset of the Instability

As the arc current is increased, the spoke appears at decreasing values of magnetic field (Fig. XII-14). To investigate this effect, the space-potential distribution was measured for various values of arc current and magnetic field (Fig. XII-15). The spoke appears when the space potential rises to approximately 10 volts with respect to the walls, and decreases for increasing radius. It is observed that the spoke rotates about the field lines in the left-handed sense, which is the direction of the Hall drift for an outward-directed electric field. The spoke is present only in the region in which the electric field is directed outward. The inner boundary of this region moves closer to the axis as the magnetic field is increased.
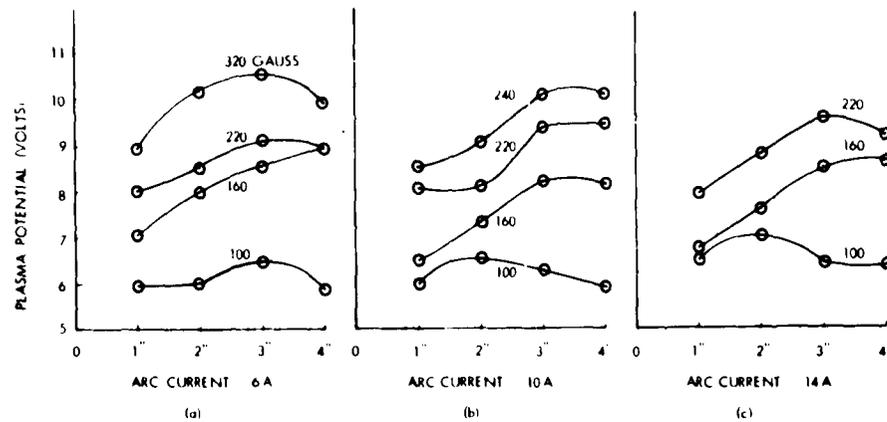
Fig. XII-15. Plasma-potential distribution for various magnetic fields and arc currents. Top curve indicates distribution as instability appears.

3. Time-Resolved Space-Potential Measurements

To correlate the motion of the spoke and particles within the spoke with theory, a knowledge of the density and potential gradients is necessary. The density measurements have been previously reported,[1] and space-potential measurements have now been completed.

Figure XII-16 shows the variation of space potential with time for various radial positions. Since the spoke moves approximately as a rigid body, this is equivalent to plotting the potential around the circumference of a circle centered on the system's axis. The bottom curve is proportional to plasma density at a distance of 3 inches from the system axis. It shows the slowly rising leading edge and the sharply dropping trailing edge of the distrubance.

If the space potential within the spoke is plotted against radius, a linear descent with radius is obtained. This explains our observation that the spoke is clearest when the magnetic field is shaped so that its axial component falls off approximately as $1/r$ to give an approximately constant value of E/rB. The rotational frequency calculated from the Hall mobility agrees with the observed behavior by better than a factor of 2, over a range of magnetic field from 300 gauss to 500 gauss.
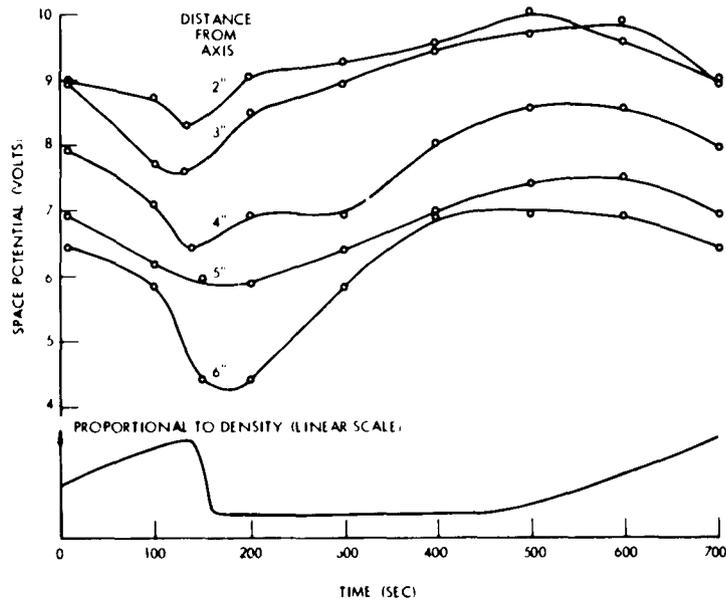
Fig. XII-16. Space potential vs time. Period of rotation, 600 μsec. Density 3 inches from axis plotted for reference. Azimuthal E-field to the right corresponds to outward radial Hall drift.

When the space potential in Fig. XII-16 descends to the right at a given radius, the result is an azimuthal electric field that drives particles radially outward at a speed determined by the Hall mobility. These phenomena occur in the region of increasing density. The Hall mobility tends to force particles inward after the spoke has passed, possibly explaining its sharp trailing edge. The net flux of particles will be outward, however, since the region of inward flow is one of very low density.

In comparing the radial flux of particles caused by the instability with the unperturbed flux, we note that the radial and azimuthal electric fields are roughly the same size, but the Hall mobility is much larger than the perpendicular mobility. With this kind of reasoning, the particle flux can be conservatively estimated to be a factor of 10 larger than the unperturbed flux.

D. L. Morse

References

1. D. L. Morse, Instability in the hollow-cathode discharge, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M.I.T., April 15, 1963, p. 54.

C. INTERACTION BETWEEN AN ELECTRON BEAM AND PLASMAS

1. Model Description

The plasma consists of electrons neutralized by ions of finite mass. Their unperturbed charge density may have a transverse variation. The dc charge density and the well-defined velocity along the axis of the waveguide of the electrons in the beam also may have a transverse variation. The electrons in the beam also are neutralized by ions of finite mass.

The waveguide is a hollow, uniform, metal tube of arbitrary cross section. Therefore we shall use a system of generalized orthogonal coordinates with the z axis taken to be parallel to the axis of the waveguide. There is a uniform, finite, axial magnetostatic field along the axis of the waveguide.

The treatment is relativistic and non quasi-static and makes use of the small-signal theory. We neglect temperature and pressure gradients, as well as collisions.

In our previous report[1] we derived the dielectric tensor, which proved to be non-Hermitian because we neglected the term $\bar{v}_o \times \bar{B}$ in the Lorentz force equation. We are not permitted to omit this term, even for a nonrelativistic treatment. The consideration of the $\bar{v}_o \times \bar{B}$ term will make the dielectric tensor Hermitian, as it must be for a lossless passive system.

2. Dielectric Tensor

From the Lorentz force equation, we find

$$\bar{v}_m = j\bar{\bar{M}}_m(\bar{E} + \bar{v}_{om} \times \bar{B}) + j\bar{i}_z \frac{\bar{v}_{mT} \cdot \nabla_T v_{mo}}{\omega_{rm}}. \tag{1}$$

The subscript m denotes the species of particle and can be e for plasma electrons, i for ions, and b for beam electrons.

The tensor $\bar{\bar{M}}_m$ has the form:

$$\bar{\bar{M}}_m = \begin{bmatrix} M_{\perp m} & -jM_{xm} & 0 \\ jM_{xm} & M_{\perp m} & 0 \\ 0 & 0 & M_{\parallel m} \end{bmatrix}. \tag{2}$$

Let $R_{om} = [1-(v_{om}^2/c^2)]^{-1/2}$; $\eta_m = e_m/m_{om}$; $\omega_{cm} = (\eta_m/R_{om})\,B_o$, and $\omega_{rm} = \omega + jy\,v_{om}$. Then the elements of the mobility tensor are

$$M_{\perp m} = \frac{\eta_m}{R_{om}}\frac{\omega_{rm}}{\left(\omega_{cm}^2 - \omega_{rm}^2\right)} \tag{3a}$$

$$M_{xm} = \frac{\eta_m}{R_{om}}\frac{\omega_{cm}}{\left(\omega_{cm}^2 - \omega_{rm}^2\right)} \tag{3b}$$

$$M_{\|m} = -\frac{\eta_m}{R_{om}^3}\frac{1}{\omega_{rm}}. \tag{3c}$$

By using the following relation from Maxwell's equations:

$$B_T = j\frac{(\nabla \times \overline{E})_T}{\omega} = -j\frac{\overline{i}_z \times [\nabla_T E_z + y\,\overline{E}_T]}{\omega}, \tag{4}$$

Eq. 1 may be written

$$\overline{v}_m - j\overline{\overline{M}}_{o\beta m}\overline{E}. \tag{5}$$

Here,

$$\overline{\overline{M}}_{op,\,m} = \overline{\overline{M}}_m \left[\begin{array}{cc:cc} \dfrac{\omega_{rm}}{\omega} & \overline{I}_T & j\dfrac{v_{om}}{\omega} & \nabla_T \\ \hdotsfor{4} \\ 0 & & 1 & \end{array} \right], \tag{6}$$

and $\overline{I}_T$ is the 2 × 2 identity matrix.

The dielectric tensor $\overline{\overline{K}}_{op}$ is defined by the equation

$$\overline{\overline{K}}_{op} \cdot \overline{E} = \overline{E} + \frac{\overline{J}}{j\omega\epsilon_o}. \tag{7}$$

By using the definition of current densities, the law of conservation of charges (Eqs. 1 and 7), we obtain

$$\overline{\overline{K}}_{op} = \overline{\overline{I}} + \frac{1}{\epsilon_o\omega}\sum_m \left[\begin{array}{c:c} \rho_{om}\overline{I}_T & 0 \\ \hdotsfor{2} \\ j\nabla_T\dfrac{\rho_{om}v_{om}}{\omega_{rm}} & \dfrac{\rho_{om}}{\omega_{rm}} \end{array} \right] \overline{\overline{M}}_{op,\,m}, \tag{8}$$

with the explicit form

$$
\bar{\bar{K}}_{op} =
\left[
\begin{array}{cc:c}
K_\perp & -jK_x & \\
& & j\dfrac{K_4}{k_o}\nabla_T - \dfrac{K_5}{k_o}\bar{i}_z \times \nabla_T \\
jK_x & K_\perp & \\
\hdotsfor{3} \\
j\nabla_T \cdot \dfrac{K_4}{k_o} + \bar{i}_z \cdot \nabla_T \times \dfrac{K_5}{k_o} & K_{||} - \nabla_T \cdot \dfrac{K_6}{k_o^2}\nabla_T + j\bar{i}_z \cdot \nabla_T \times \dfrac{K_7}{k_o^2}\nabla_T &
\end{array}
\right] . \tag{9}
$$

It should be noted that in the tensor of Eq. 9, which determines the electric displacement vector $\bar{D} = \bar{\bar{K}} \cdot \bar{E}$, the upper right operator is a 1×2 matrix operating on E_z, and the lower left operator is a 2×1 matrix operating on the transverse components of $\bar{E}$; the lower right operator is a 1×1 matrix operating on only E_z.

Let

$$
\omega_{pmT}^2 = \frac{\eta_m}{R_{om}}\frac{\rho_{om}}{\epsilon_o} = \frac{\rho_{om}e_m}{\epsilon_o m_{mT}}
$$

and

$$
\omega_{pmz}^2 = \frac{\eta_m}{R_{om}^3}\frac{\rho_{om}}{\epsilon_o} = \frac{\rho_{om}e_m}{\epsilon_o m_{mz}},
$$

Then the elements of the $\bar{\bar{K}}_{op}$ tensor are given by

$$
K_\perp = 1 + \frac{1}{\omega^2}\sum_m \frac{\omega_{pmT}^2\omega_{rm}^2}{\omega_{cm}^2 - \omega_{rm}^2} \tag{10a}
$$

$$
K_x = \frac{1}{\omega^2}\sum_m \frac{\omega_{pmT}^2\omega_{cm}\omega_{rm}}{\omega_{cm}^2 - \omega_{rm}^2} \tag{10b}
$$

$$
K_{||} = 1 - \sum_m \frac{\omega_{pmz}^2}{\omega_{rm}^2} \tag{10c}
$$

$$
\frac{K_4}{k_o} = \frac{1}{\omega^2}\sum_m \frac{\omega_{pmT}^2\omega_{rm}}{\omega_{cm}^2 - \omega_{rm}^2}V_{om} \tag{10d}
$$

$$\frac{K_5}{k_o} = \frac{1}{\omega^2} \sum_m \frac{\omega_{pm}^2 T^{\omega}_{cm}}{\omega_{cm}^2 - \omega_{rm}^2} \, v_{om} \tag{10e}$$

$$\frac{K_6}{k_o^2} = \frac{1}{\omega^2} \sum_m \frac{\omega_p^2 \, \omega_{pm}^2 T}{\omega_{cm}^2 - \omega_{rm}^2} \, v_{om}^2 \tag{10f}$$

$$\frac{K_7}{k_o^2} = \frac{1}{\omega^2} \sum_m \frac{\omega_{pm}^2 T^{\omega}_{cm}}{\omega_{rm} \left(\omega_{cm}^2 - \omega_{rm}^2 \right)} \, v_{om}^2 \tag{10g}$$

3. Hermitian Character of the $\overline{\overline{K}}_{op}$ Tensor

In the following discussion we shall write $\overline{\overline{K}}_{op}$ as $\overline{\overline{K}}$. For propagating wave, $\gamma = j\beta$, in a lossless passive system we can show[2] that

$$\int_A \overline{E}_2^* \cdot \overline{\overline{K}}_h \cdot \overline{E}_1 \, da - \int_A \overline{E}_1 \cdot \overline{\overline{K}}_h^* \cdot \overline{E}_2^* \, da = 0. \tag{11}$$

Relation (11) applied for $\overline{E}_1 = \overline{E}_2 = \overline{E}$ shows that $\int_A \overline{E}^* \cdot \overline{\overline{K}} \cdot \overline{E} \, da$ is a real number. We may easily see

$$\frac{1}{2} \nabla \cdot (\overline{E} \times \overline{H}^*) = -j2\omega \left(\frac{1}{4} \mu_o | \overline{H} |^2 - \frac{1}{4} \epsilon_o \overline{E}^* \cdot \overline{\overline{K}} \cdot \overline{E} \right) \tag{12}$$

that for a lossless passive system Re $\int_A \frac{1}{2} (\overline{E}_T \times \overline{H}_T^*) \cdot \overline{i}_z \, da = 0$. The integral $\int_A \overline{E}^* \cdot \overline{\overline{K}} \cdot \overline{E} \, da$ must be a real number.

In general, $\gamma = \alpha + j\beta$. In this case, the $\overline{\overline{K}}$ tensor has Hermitian and anti-Hermitian parts: $\overline{\overline{K}} = \overline{\overline{K}}_h + \overline{\overline{K}}_a$. The definition for $\overline{\overline{K}}_a$ is

$$\int \overline{E}_2^* \cdot \overline{\overline{K}}_a \cdot \overline{E}_1 \, da + \int \overline{E}_1 \cdot \overline{\overline{K}}_a^* \cdot \overline{E}_2^* \, da = 0. \tag{13}$$

In the above-given proof for the Hermitian character of $\overline{\overline{K}}$ for $\gamma = j\beta$, we have used the property that the K's in Eqs. 10 are real for $\gamma = j\beta$. Therefore for $\gamma = \alpha + j\beta$ the $\overline{\overline{K}}_h$ will be found when we consider the real part of the K's, and the $\overline{\overline{K}}_a$, when we consider the imaginary part of the K's.

There are many applications of Eq. 12. One of them is the extension of the variational principle for the propagation constant γ to the beam-plasma system.[2]

4. Plane Waves

If we assume a plane wave of the form $\exp(j\omega t - j\overline{k} \cdot \overline{r})$ with $\overline{k}$ on the x-z plane, the dielectric tensor, Eq. 9, becomes

$$K = \begin{bmatrix} K_\perp & \vdots & -jK_x & \vdots & +K_4 \dfrac{k_x}{k_o} \\ \cdots & & \cdots & & \cdots \\ jK_x & \vdots & K_\perp & \vdots & jK_5 \dfrac{k_x}{k_o} \\ \cdots & & \cdots & & \cdots \\ K_4 \dfrac{k_x}{k_o} & \vdots & -jK_5 \dfrac{k_x}{k_o} & \vdots & K_\| + K_6 \dfrac{k_x^2}{k_o^2} \end{bmatrix} . \tag{14}$$

This relation has already been used for the investigation of various waves.[3]

5. Field Analysis

We shall assume hereafter uniform dc charge density and dc beam velocity. The longitudinal fields are coupled, as in the case of plasma alone[4,5]:

$$\nabla_T^2 E_z + a E_z = b H_z \tag{15}$$

$$\nabla_T^2 H_z + c H_z = d E_z . \tag{16}$$

The coefficients now are

$$a = \frac{K_\| \left(K_\perp k_o^2 + \gamma^2 \right)}{K_\perp - j \dfrac{2\gamma k_4}{k_o} - \dfrac{\gamma^2 K_6}{k_o^2} + K_4^2 - K_\perp K_6} \tag{17a}$$

$$b = j\omega\mu_o \frac{j\gamma \left(K_x - \dfrac{j\gamma K_5}{k_o} \right) + k_o (K_\perp K_5 - K_x K_4)}{K_\perp - j \dfrac{2\gamma K_4}{k_o} - \dfrac{\gamma^2 K_6}{k_o^2} + K_4^2 - K_\perp K_6} \tag{17b}$$

$$c = \gamma^2 + k_o^2 K_\perp - k_o^2 K_x^2 \frac{\left(1 - \dfrac{j\gamma}{k_o}\dfrac{K_5}{K_x} \right)^2 - K_\perp \dfrac{K_5^2}{K_x^2} - K_6 + 2 \dfrac{K_4 K_5}{K_x}}{K_\perp - j \dfrac{2\gamma K_4}{k_o} - \dfrac{\gamma^2 K_6}{k_o^2} + K_4^2 - K_\perp K_6} \tag{17c}$$

$$d = -\frac{\epsilon_o}{\mu_o} K_\| b . \tag{17d}$$

The transverse fields are given in terms of the longitudinal fields by the relations

$$
\begin{bmatrix} \mathbf{E}_T \\ \overline{\mathbf{H}}_T \\ \mathbf{i}_z \times \overline{\mathbf{E}}_T \\ \mathbf{i}_z \times \overline{\mathbf{H}}_T \end{bmatrix} = \begin{bmatrix} P & r & Q & s \\ T & p & U & q \\ -Q & -s & P & s \\ -U & -q & T & p \end{bmatrix} \begin{bmatrix} \nabla_T \mathbf{E}_z \\ \nabla_T \mathbf{H}_z \\ \mathbf{i}_z \times \nabla_T \mathbf{E}_z \\ \mathbf{i}_z \times \nabla_T \mathbf{H}_z \end{bmatrix} .
\tag{18}
$$

Here,

$$
P = p - \omega\epsilon_0 \frac{K_4}{k_0} s - j\omega\epsilon_0 \frac{K_5}{k_0} r
\tag{19a}
$$

$$
Q = q + \omega\epsilon_0 \frac{K_4}{k_0} r - j\omega\epsilon_0 \frac{K_5}{k_0} s
\tag{19b}
$$

$$
T = t - \omega\epsilon_0 \frac{K_4}{k_0} q - j\omega\epsilon_0 \frac{K_5}{k_0} p
\tag{19c}
$$

$$
U = u + \omega\epsilon_0 \frac{K_4}{k_0} p - j\omega\epsilon_0 \frac{K_5}{k_0} q,
\tag{19d}
$$

and p, q, r, s, t, and u are the known expressions.[5]

Our formulation for the beam-plasma system is similar to the formulation for plasma alone. Hence, for the rest of this report, we can use techniques and results that are already known.[4,6]

The dispersion relation is

$$
p^4 - (a+c)p^2 + ac - bd = 0.
\tag{20}
$$

The determinantal equation is derived from boundary conditions. For a completely filled circular waveguide it has the same form as for an anisotropic plasma.

$$
A_1 \frac{J_n'(p_1 r_0)}{J_n(p_1 r_0)} - A_2 \frac{J_n'(p_2 r_0)}{J_n(p_2 r_0)} = jn A_3,
\tag{21}
$$

where A_1, A_2, and A_3 are algebraic functions of ω, γ, v_{om}, ω_{pm}, ω_{cm}, p_1, and p_2.

6. Energy-Power Theorems and Variation Theorems

Assuming that the fields are of the form $\overline{A}(u_1, u_2, z, t) = \overline{A}(u_1, u_2) \exp(j\omega t - \gamma z)$ and manipulating Maxwell's equations, we obtain

$$aP_e = -\omega \ \text{Im} \ U_{\epsilon a} \tag{22}$$

$$aQ_e = \omega (U_m - U_{\epsilon h}) \tag{23}$$

$$\beta P_e = \omega (U_{mT} - U_{mz} + U_{\epsilon Th} - U_{\epsilon zh} + 2 \ \text{Re} \ U_{\epsilon zTa}) \tag{24}$$

$$\beta Q_e = \omega \ \text{Im} \ (U_{\epsilon za} - U_{\epsilon Ta} - 2U_{\epsilon zTh}), \tag{25}$$

where

$$P_e + jQ_e = \int_A \frac{1}{2} \bar{E}_T \times \bar{H}_T^* \cdot \bar{i}_z \ da \tag{26a}$$

$$U_\epsilon = \int \frac{1}{4} \epsilon_o \bar{E}^* \cdot \bar{\bar{K}} \cdot \bar{E} \ da \tag{26b}$$

$$U_{\epsilon T} = \int \frac{1}{4} \epsilon_o \bar{E}_T^* \cdot \bar{\bar{K}}_T \cdot \bar{E}_T \ da \tag{26c}$$

$$U_{\epsilon z} = \int \frac{1}{4} \epsilon_o E_z^* K_z E_z \ da \tag{26d}$$

$$U_{\epsilon zT} = \int \frac{1}{4} \epsilon_o \bar{E}_T^* \cdot \bar{\bar{K}}_{Tz} E_z \ da \tag{26e}$$

$$U_m = \int \frac{1}{4} \mu_o |\bar{H}|^2 \ da. \tag{26f}$$

The subscript h (or a) means that we must consider the Hermitian (or anti-Hermitian) part of $\bar{\bar{K}}$ in the corresponding formulae.

By use of Maxwell's equations and their variations, we get

$$\delta(aP_e) = - \delta(\omega \ \text{Im} \ U_{\epsilon a}) \tag{27}$$

$$\delta aQ_e = -\frac{P}{2} \ \text{Re} \ \int (\delta \bar{E}_T \times \bar{H}_T^* - \bar{E}_T \times \delta \bar{H}_T^*) \cdot \bar{i}_z \ da$$

$$+ \delta \omega (U_{mT} - U_{mz}) + \int \left[\frac{1}{4} \epsilon_o E_z^* \delta(\omega K_{zh}) E_z - \frac{1}{4} \epsilon_o \bar{E}_T^* \cdot \delta(\omega \bar{\bar{K}}_{Th}) \cdot \bar{E}_T \right] da$$

$$- \frac{\omega \epsilon_o}{2} \ \text{Re} \ \int [\delta E_Z^* (\bar{\bar{K}}_{zTh} \cdot \bar{E}_T + K_{za} E_z) - \delta \bar{E}_T^* \cdot (\bar{\bar{K}}_{Tzh} \cdot E_z + \bar{\bar{K}}_{Ta} \cdot \bar{E}_T)] da$$

$$- \frac{\epsilon_o}{2} \text{Re} \ \int \hat{\bar{E}}_T^* \cdot \delta(\omega \bar{\bar{K}}_{aTz}) E_z \ da \tag{28}$$

$$\delta\beta P_e = -\frac{\alpha}{2} \text{ Im } \int (\delta\overline{E}_T \times \overline{\overline{H}}_T^* - \overline{E}_T \times \overline{\delta H}_T^*) \cdot i_z \, da + U_m \delta\omega$$

$$+ \int \frac{1}{4} \epsilon_o \overline{E}^* \cdot \delta(\omega\overline{\overline{K}}_h) \cdot \overline{\overline{E}} \, da - \frac{\omega\epsilon_o}{2} \text{ Re } \int \delta\overline{E}^* \cdot \overline{\overline{K}}_a \cdot \overline{E} \, da \qquad (29)$$

$$\delta(\beta Q_e) = \delta\left[\omega \text{ Im } (U_{\epsilon za} - U_{\epsilon TQ} - 2U_{\epsilon zTh})\right] \qquad (30)$$

If $\alpha = 0 \ (\overline{\overline{K}} = \overline{\overline{K}}_h)$, Eq. 29 gives

$$\frac{\partial\omega}{\partial\beta} \simeq \frac{P_e - \int \frac{1}{4}\epsilon_o \overline{E}^* \cdot \dfrac{\partial(\omega\overline{\overline{K}})}{\partial\beta} \cdot \overline{E} \, da}{\int \left[\frac{1}{4}\mu_o |\overline{H}|^2 + \frac{1}{4}\epsilon_o \overline{E}^* \cdot \dfrac{\partial(\omega\overline{\overline{K}})}{\partial\omega} \cdot \overline{E}\right] da} \equiv \frac{\langle P_e \rangle + \langle P_m \rangle}{\langle W \rangle}. \qquad (31)$$

Therefore Eq. 31 is applicable even when $\overline{\overline{K}}$ is an operator.

For $\alpha \neq 0$ we may derive the expression for the average power density (which is unique[7]) from Eq. 22. Equation 22 is valid for any $\overline{\overline{K}}$ tensor. When our system is lossless and a conservation principle for the time-averaged power may be derived for it, we have

$$\alpha P_e = -\alpha P_M. \qquad (32)$$

By use of Eq. 22, we get

$$P_M' = \frac{\omega}{\alpha} \text{ Im } U_{\epsilon a}. \qquad (33)$$

Equation 33 is valid for $\alpha \neq 0$. For $\alpha \rightarrow 0$ we find

$$\lim_{\alpha \rightarrow 0} P_M = - \int \frac{1}{4}\epsilon_o \overline{E}^* \cdot \frac{\partial(\omega\overline{\overline{K}}_h)}{\partial} \cdot \overline{E} \, da,$$

which checks with the well-known result for $\alpha = 0$.[6]

<div align="right">P. E. Serafim, A. Bers</div>

References

1. P. E. Serafim, Interaction between an electron beam and plasma, Quarterly Progress Report No. 64, Research Laboratory of Electronics, M.I.T., January 15, 1962, pp. 95-103.

2. A. Bers, Electromagnetic Waves in Passive and Active Systems, Course 6.637 Lecture, M.I.T., Spring 1962, pp. 38-43.

3. R. J. Briggs and A. Bers, Transverse free waves on electron beams, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M.I.T., October 15, 1962, pp. 35-43.

4. A. A. Th. M. Van Trier, Appl. Sci. Research B3, 305 (1953).

5. W. P. Allis, S. J. Buchsbaum, and A. Bers, Waves in Anisotropic Plasmas (The M.I.T. Press, Cambridge, Mass., 1963), pp. 135-136, give these expressions as upper case letters.

6. W. P. Allis, S. J. Buchsbaum and A. Bers, op. cit.

7. D. L. Bobroff, H. A. Haus, and J. W. Klüver, On the small signal power theorem of electron beams, J. Appl. Phys. 33, 2932-2942 (1962).

D. INTERACTION OF AN ELECTRON BEAM WITH IONS IN A WARM PLASMA OF FINITE TRANSVERSE DIMENSIONS

It has been shown that a one-dimensional model predicts a strong interaction of an electron beam with the ions of a plasma when the plasma electrons are sufficiently warm.[1] The interaction is of the reactive-medium type, and it occurs when the beam space-charge wavelength is less than the plasma Debye wavelength. The present report is concerned with the extension of these results to a beam-plasma system with finite transverse dimensions.
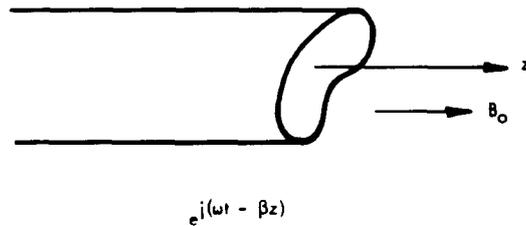
Fig. XII-17. Beam-plasma system.

It is assumed that both the beam and the plasma are homogeneous, and fill a cylindrical waveguide structure (Fig. XII-17). The beam moves with unperturbed velocity, v_o, parallel to a large magnetic field, B_o, which constrains electrons to motion along the field lines. The ions are assumed to be cold; however, transverse motion of the ions is allowed. When the quasi-static assumption is made, it is found that the potential satisfies the two-dimensional Helmholtz equation

$$\nabla_T^2 \phi + p^2 \phi = 0,$$ (1)

where ∇_T^2 is the Laplacian operator in the transverse plane. Also

$$p^2 = -\beta^2 \frac{K_{\parallel}}{K_{\perp}},$$ (2)

where

$$K_{\parallel} = 1 - \frac{\omega_{pi}^2}{\omega^2} - \omega_{pe}^2 \int \frac{f_o(v_z)\, dv_z}{(\omega-\beta v_z)^2} - \frac{\omega_{pb}^2}{(\omega-\beta v_o)^2}$$ (3)

$$K_\perp = 1 - \frac{\omega_{pi}^2}{\omega^2 - \omega_{ci}^2}$$ (4)

and all symbols have been defined previously.[1]

As in the one-dimensional case, strong interaction with ions is expected only when
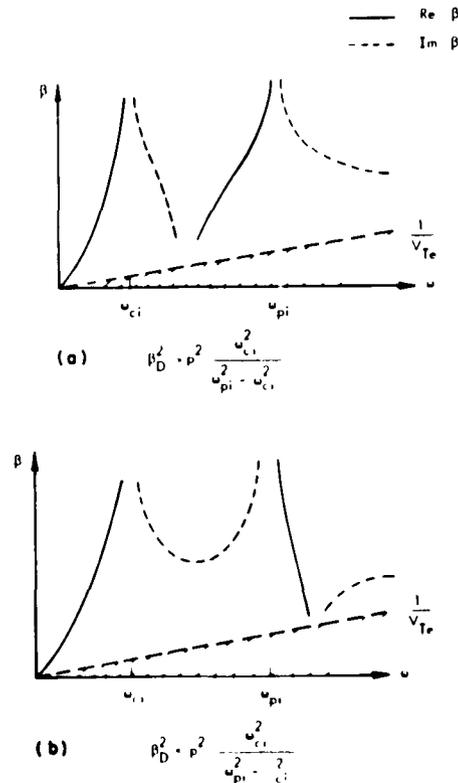
Fig. XII-18. Dispersion in a hot electron plasma.

$v_o \ll V_{Te}$, where V_{Te} is the average thermal velocity of the plasma electrons. Therefore, attention will be restricted to the range in which $\omega \ll \beta V_{Te}$. In this limit,

$$K_\parallel \approx 1 - \frac{\omega_{pi}^2}{\omega^2} + \frac{\omega_{pe}^2}{\beta^2 V_{Te}^2} - \frac{\omega_{pb}^2}{(\omega - \beta v_o)^2}.$$ (5)

For the plasma in the absence of the beam, the dispersion relation can be written as

$$\beta^2 = -\frac{\beta_D^2 + p^2 K_\perp}{1 - \frac{\omega_{pi}^2}{\omega^2}},$$

(6)

where $\beta_D = \omega_{pe}/V_{Te}$ is the Debye wave number. This dispersion is illustrated in Fig. XII-18 for $\omega_{ci} < \omega_{pi}$. In the limit $V_{Te} \to \infty$, the dispersion becomes identical to that of a cold ion cloud, as would be expected.[2] Normally, however, for reasonable electron temperatures, the resonance at ω_{pi} belongs to a forward wave.

The dispersion equation for the beam-plasma system is a 4th-order equation in β.
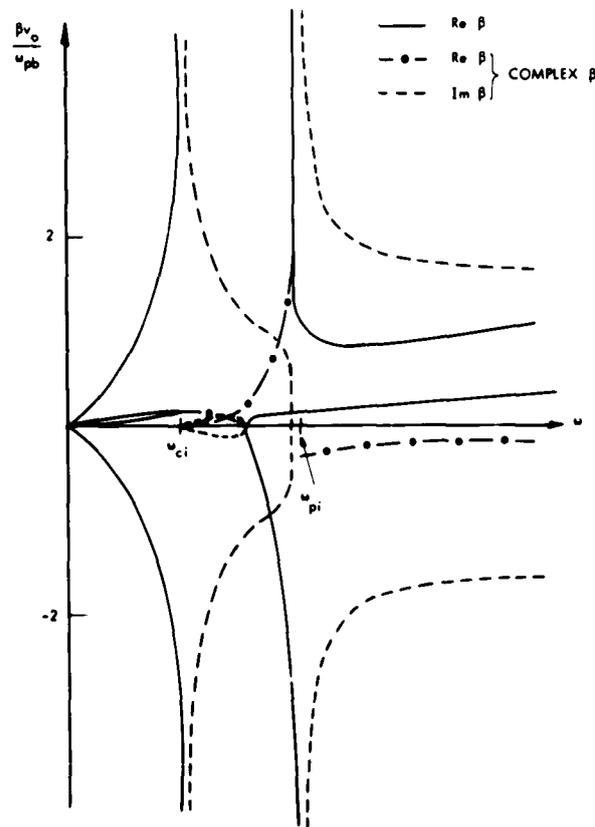
Fig. XII-19. Dispersion for $\dfrac{n_b}{n_p} \dfrac{T_e}{2V_o} = 0.5$.

The results of some computations are shown in Figs. XII-19, XII-20, XII-21 for protons with $\beta_{pb}^2 = p^2$, $n_b = 10^{-2} n_p$, and $\omega_{pi} = 2\omega_{ci}$, and for values of the "temperature"
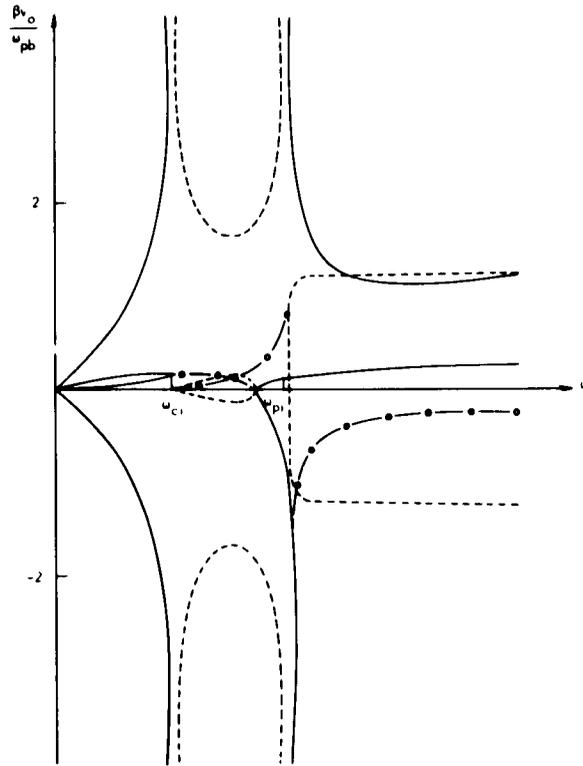
Figure XII-20.

Dispersion for $\dfrac{n_b}{n_p} \dfrac{T_e}{2V_o} = 1$.
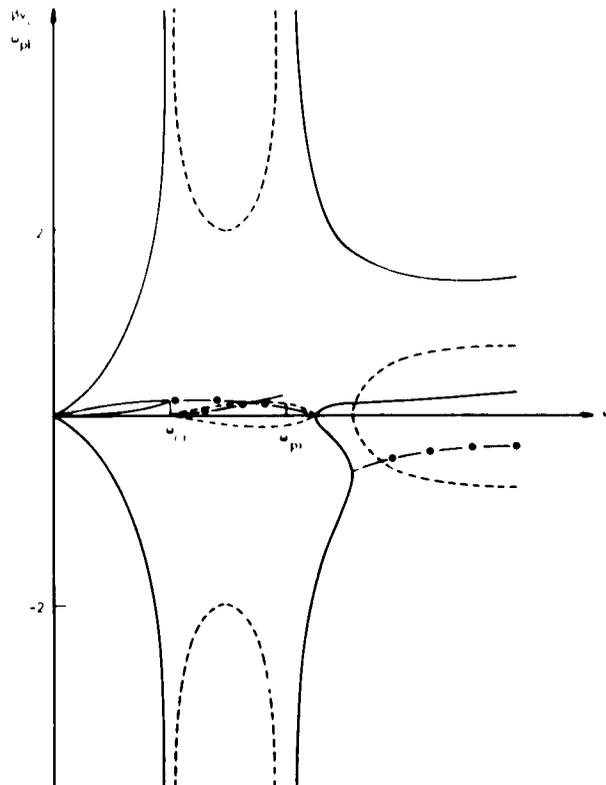
Figure XII-21.

Dispersion for $T_e = \infty$.

parameter $\dfrac{n_b}{np}\dfrac{T_e}{2V_o}=\dfrac{\lambda_D^2}{\lambda_{pb}^2}$ equal to 0.5, 1, and ∞. The condition for obtaining real β for

$\omega > \omega_{pi}$ now is

$$\beta_{pb}^2 + p^2 \frac{\omega_{ci}^2}{\omega_{pi}^2 - \omega_{ci}^2} > \beta_D^2, \tag{7}$$

in agreement with the numerical computations.

The interpretation of these plots by means of the amplifying criterion that was recently derived by the authors is being investigated. It is interesting to note that the complex root of β with Im $\beta \to \infty$ at $\omega = \omega_{pi}$ carries negative kinetic power if the Debye wave number in the range

$$\beta_{pb}^2 + p^2 \frac{\omega_{ci}^2}{\omega_{pi}^2 - \omega_{ci}^2} > \beta_D^2 > p^2 \frac{\omega_{ci}^2}{\omega_{pi}^2 - \omega_{ci}^2} . \tag{8}$$

This raises a strong suspicion that this wave is an amplifying wave with a spatial growth rate tending to infinity, under the condition of Eq. 8.

R. J. Briggs, A. Bers

References

1. R. J. Briggs and A. Bers, Electron-beam interactions with ions in a warm plasma, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M.I.T., April 15, 1963, pp. 65-73.

2. L. D. Smullin and P. Chorney, Proc. International Symposium on Electronic Waveguides (Polytechnic Institute of Brooklyn, New York, 1958), pp. 229-247.

E. TEST AND MODIFICATION OF LARGE SUPERCONDUCTING SOLENOID

The large superconducting solenoid described at length in previous reports[1-4] has been completed and temperature-cycled twice. The coils could not be cooled below the superconducting transition temperature; the system has a repairable thermal short.

To give an idea of what happened, we summarize the pertinent assembly details. Figure XII-22 shows a cross section of one end of the magnet. The view is not drawn to scale so that the critical small dimensions will be visible. Unlabeled dimensions have no relevance to the present discussion. Some vacuum flanges and other detail in irrelevant places are not shown. All radiation shields are held off surfaces by nylon studs; bolt and screw locations are shown by simple lines, as pins.

The 0.25-in. span between magnet coils and shield H was originally intended to be larger (≥ 0.5 in.), but the individual magnet coils (24 of them) spread axially when wound, and thus reduced the clearance and created part of our problem.

As the coils cool, the supporting column shrinks almost 0.25 in.; the spacings shown are for the system "cold." The system was aligned before evacuation and cooling. During cooling, however, the supporting columns shrank unevenly, moving the nitrogen shield to the left, and the magnet coils to the right. Thus the magnet coil structure touched the shield H; also, the coils swung (out of the paper in the sketch) and a metallic connection was made between the coils and the side of the nitrogen shield F. Total motion was approximately 0.25 in. The coils cooled to approximately 15°K (estimated) through circulation of helium through the pipes, K, attached to the coil. By forcing excess helium through the system, the thermal switches at point J could be made superconducting
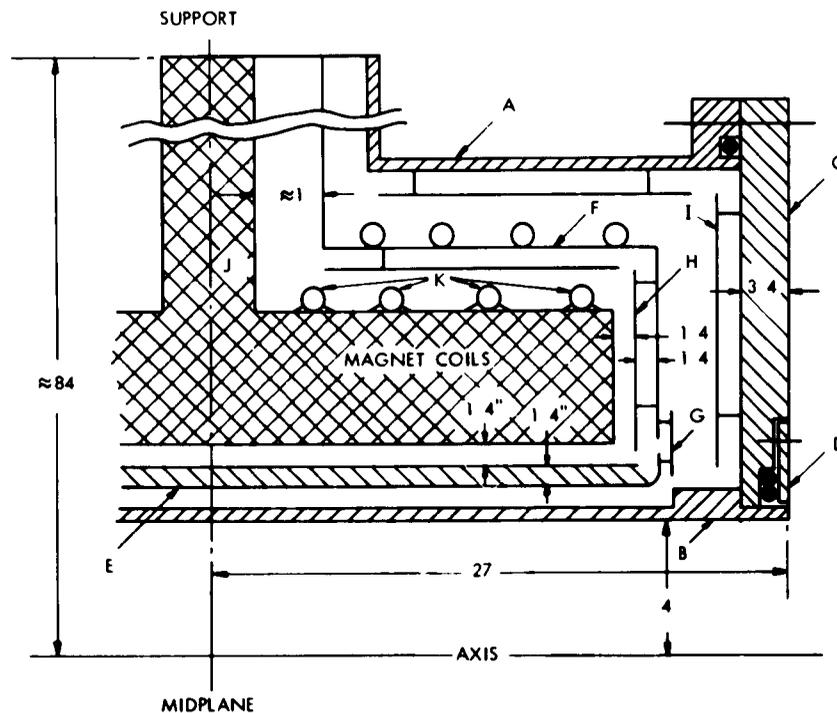
Fig. XII-22. Sketch (not to scale) of cross section of magnet, upper right quadrant, showing critical dimensions, in inches.
A. Cylindrical vacuum wall, O. D.
B. Cylindrical vacuum wall, I. D.
C. End flange
D. Sealing ring and O-rings, I. D.
E. Nitrogen shield, including cooling pipes and radiation shield, I. D.
F. Nitrogen shield, O. D. and end flanges
G. Annular joining ring, nitrogen shield
H. Radiation shield, supported off end of nitrogen shield
I. Radiation shield, supported off end of vacuum wall
J. Position of thermal switches
K. Helium precooling line
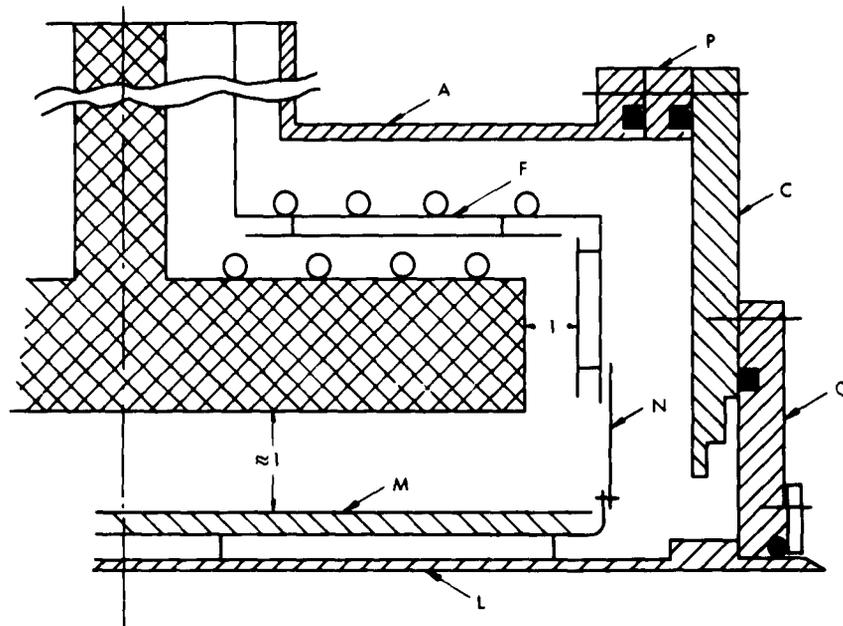
Fig. XII-23. Modification to magnet.
  L. New vacuum wall, I. D.
  M. New nitrogen shield, I. D.
  N. Thermal sealing ring
  P. Axial spacer
  Q. End flange adaptor

because the helium line ends at a reservoir there.

No contact between the nitrogen shield and the outer wall took place. For the coils and nitrogen shield together, the thermal isolation is good, even at a pressure of 2×10^{-5} mm Hg. Temperature rise was 10-15°K/day at 77°K.

Heat transfer from the cooling pipes K to the magnet coils has been recalculated and found more than adequate for initial solenoid tests. The largest thermal impedance lies in the coils themselves, where the wire is insulated by 0.001-in. nylon and 0.010-in. mylar between wire layers. The cooling time constant is ~100 sec at 20°K, and will be higher near 4°K. Times of 100-1000 sec are acceptable. It is calculated that the thermal contacts that had been found could reasonably account for the 25-watt heat input to the magnet coil (as determined by helium enthalpy rate into and out of the magnet coil structure).

A relatively simple modification is planned to correct the difficulty at the cost of 1 inch in working radius. The modification is sketched in Fig. XII-23 (not drawn to scale). The vacuum cylinder A, the end plate C, and the nitrogen shield F are retained. These are the most expensive items. A new I. D. vacuum wall L, of 3-in. radius, has been made, and a new inner nitrogen shield M is attached to it with nylon studs. Cooling is brought to it by a pipe and bellows from shield component F. Radial clearance to the

magnet coil I. D. is now almost 1 inch. Also, the end of shield F is cut off, and a cylin-
drical strip is inserted to extend F axially 0.75 inch. The nitrogen shielding is com-
pleted by a new annular copper ring N, bolted to M, and loosely pinned to the end wall F.

The axial modification requires a ring spacer P on the outside diameter, and the
radial modification requires plate Q, which clamps to C with an O-ring and holds the I. D.
wall L. These modifications have been made at both ends of the magnet.

Dimensions at the outer diameters (magnet coils radially to F, for example) are large
and require no modification.

D. J. Rose, L. M. Lidsky, E. Thompson, J. Woo

References

1. D. J. Rose and L. J. Donadieu, Quarterly Progress Report No. 62, Research
Laboratory of Electronics, M. I. T., July 15, 1961, p. 68.

2. L. J. Donadieu and D. J. Rose in Proc. International Conference on High Magnetic
Fields, edited by H. Kolm (The M. I. T. Press, Cambridge, Mass., 1961), pp. 358-369.

3. L. J. Donadieu, Large-volume superconducting solenoid, Quarterly Progress
Report No. 66, Research Laboratory of Electronics, M. I. T., July 15, 1962, p. 139.

4. L. J. Donadieu, Superconducting solenoid, Quarterly Progress Report No. 68,
Research Laboratory of Electronics, M. I. T., January 15, 1963, p. 86.

F. ORBIT STABILITY IN THE CORKSCREW

1. Introduction

The transfer of longitudinal to transverse kinetic energy in a helically perturbed
magnetic field has been demonstrated experimentally by Wingerson,[1] Dreicer et al.,[2]
and discussed in detail by Wingerson, Dupree, and Rose.[3] The energy transfer depends
on a resonance of the position and velocity of the particle with the orientation and mag-
nitude of the perturbing magnetic field. It has been shown,[3] and is now shown in greater
detail in this report, that this resonance is stable to first order, that is, a particle dis-
placed from the stable orbit will oscillate about the position of stability. We address
ourselves here to the question of second-order stability — do the oscillations about the
stable orbit grow or decay?

It will be shown in this report that the linearized equations of motion predict growth
of the oscillations in an axially decelerating corkscrew and decay of the oscillations in
an axially accelerating corkscrew. The equations of motion for perturbations about the
stable orbit in an optimized (in the sense explained elsewhere[3]) corkscrew are displayed.
Numerical solutions of these equations are presented, and the results discussed.

2. Linearized Equations of Motion

For the coordinate system of Fig. XII-24, the orbital equations for v, the azimuthal
velocity, and χ, the relative azimuthal angle, for total energy E_o, are

$$\frac{dv}{dz} = \omega_r(r, z) \cos \chi \tag{1}$$

$$\frac{d\chi}{dz} = -\frac{2\pi}{P(z)} + \frac{\omega_z}{[U-v^2]^{1/2}}, \tag{2}$$

where $\omega_r = qB_r/m$, $\omega_z = qB_z/m$, $U = 2E_0/m$, and $P(z)$ is the local pitch length. The assumption is made that the radial velocity, but not the radial position, can be ignored.
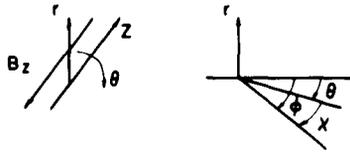
Fig. XII-24. Coordinate system for derivation of the orbital equations.

θ = ANGULAR POSITION OF FIELD MAXIMUM

ϕ = ANGULAR POSITION OF PARTICLE

If $v = v_0 + v_1$ and $\chi = \chi_0 + \chi_1$, where $v_0(z)$ and $\chi_0(z)$ refer to the unperturbed orbit of the particle, then to first order in the perturbation variables

$$\frac{dv_1}{dz} = -\omega_r(r, z) \sin \chi_0 \chi_1 \tag{3}$$

$$\frac{d\chi_1}{dz} = -\frac{2\pi}{P(z)} + \frac{\omega_z v_0 v_1}{\left[U-v_0^2\right]^{3/2}}. \tag{4}$$

Equations 3 and 4 yield a second-order differential equation for χ,

$$\frac{d^2\chi_1}{dz^2} + g(z) \frac{d\chi_1}{dz} + h(z) \chi_1 = 0, \tag{5}$$

where

$$g(z) = \frac{3}{P} \frac{dP}{dz} - \frac{1}{v_0} \frac{dv_0}{dz} \tag{6}$$

and

$$h(z) = \frac{(2\pi)^3 v_0(z) \omega_r(r, z) \sin \chi_0}{\omega_z^2 P^3(z)}. \tag{7}$$

The coefficient of χ_1 is positive for both accelerating and decelerating corkscrews. For deceleration with $\omega_r > 0$, Eq. 1 limits the particle position to the first or fourth quadrant (that is, the azimuthal velocity must increase with z). A particle in the first quadrant displaced forward in χ will be decelerated less strongly, move with too high z velocity relative to the local pitch and so move back toward the original position. Similarly, a particle displaced backward in χ will suffer stronger deceleration and move toward the original position. On the other hand, a particle in the fourth quadrant will move in the direction of the displacement, and rapidly fall out of synchronism. An accelerating corkscrew, with χ_0 necessarily in either the second or third quadrant, is stable (in the sense that small displacements lead to oscillatory motion about the origin of the displacement) only in the second quadrant. By the same arguments, if the direction of the perturbing field is reversed, then χ_0 must be in the third or fourth quadrant for oscillatory stability. In any of these cases, h(z) is positive.

Because h(z) is a monotonically varying, always positive function of z, the stability of the oscillations depends only on the sign of g(z). Equation 6 shows g(z) to be negative (growing oscillations) for decelerating corkscrews, and to be positive (decaying oscillations) for accelerating corkscrews.

3. Optimum Corkscrew

Wingerson, Dupree, and Rose demonstrate that the scattering losses for trapped particles are reduced for a corkscrew in which $dv_\perp/dz \to 0$ at both ends. We shall consider first-pass particle motions in such a system. Particularly, we demand that the perpendicular velocity in the unperturbed orbit, $v_{\perp o}$, be given by

$$v_{\perp o} = \alpha v_o \sin^2\left(\frac{\pi z}{2L}\right),$$ (8)

where α is a parameter describing the total change in perpendicular velocity, v_o is the TOTAL (vector) velocity of the particle, and the corkscrew is assumed to be of length L. The resonance condition gives

$$P(z) = \frac{2\pi v_o}{\omega_z} \left[1 - \alpha^2 \sin^4\left(\frac{\pi z}{2L}\right)\right]^{1/2}$$ (9a)

$$= P_o\left[1 - \alpha^2 \sin^4\left(\frac{\pi z}{2L}\right)\right]^{1/2}.$$ (9b)

Equations 1, 8, and 9b combine to fix the necessary variation of $\omega_r(r, t)$:

$$\omega_r(r, z) = \frac{\left(\frac{\pi \alpha}{2L}\right) \sin\left(\frac{\pi z}{L}\right)}{G(\chi_o, z) \cos \chi_o}\left\{I_o\left[\frac{2\pi r}{P(x)}\right] + I_2\left[\frac{2\pi r}{P(z)}\right]\right\},$$ (10)

where

$$G(r_0, z) = G\left(\frac{v_{\perp o}}{\omega_z}, z\right) = I_0\left[\frac{2\pi v_{\perp o}(z)}{\omega_z P(z)}\right] + I_2\left[\frac{2\pi v_{\perp o}(z)}{\omega_z P(z)}\right].$$ (11)

The algebra is simplified by introducing the new variables $x = \pi z/2L$, $v = v_\perp/v_0$, and $p(x) = P(x) = P_0$. In this notation,

$$\frac{dv}{dx} = \frac{a \sin 2x}{G(v_{\perp o}, x) \cos \chi_o} \cdot G(v, x) \cos \chi$$ (12)

and

$$\frac{d\chi}{dx} = \frac{4L}{P_0}\left[\frac{1}{(1-v^2)^{1/2}} - \frac{1}{p(x)}\right].$$ (13)

We take advantage of the fact that the unperturbed orbit is known and rewrite the equations in terms of deviations from this orbit; that is, for $\chi_1(x) = \chi(z) - \chi_0$ and $v_1(z) = v(z) - v_0(z)$.

$$\frac{dv_1}{dx} = a \sin 2x\left\{\frac{G(v_0+v_1, x)}{G(v_0, x)}(\cos\chi_1 - \tan\chi_0 \sin\chi_1) - 1\right\}$$ (14)

$$\frac{d\chi_1}{dx} = \frac{\gamma}{p(x)}\left\{\frac{1}{1 - \left[\frac{2a \sin^2 xv_1 + v_1^2}{p^2(x)}\right]^{1/2}} - 1\right\}.$$ (15)

where

$$p(x) = [1 - a_2 \sin^4(x)]$$ (16)

$$G(v_0+v_1, x) = I_0\left[\frac{a \sin^2(x) + v_1}{p(x)}\right] + I_2\left[\frac{a \sin^2(x) + v_1}{p(x)}\right]$$ (17)

$$\gamma = 4L/P_0.$$ (18)

Except for the slowly varying ratio of the G's, Eqs. 14 and 15 are first order in deviations from equilibrium. The oscillations about the unperturbed orbit are thus separated from the equations of the orbit itself, and can be studied with far greater accuracy. Notice also that no additional approximations have been made (that is, these equations are not linearized).
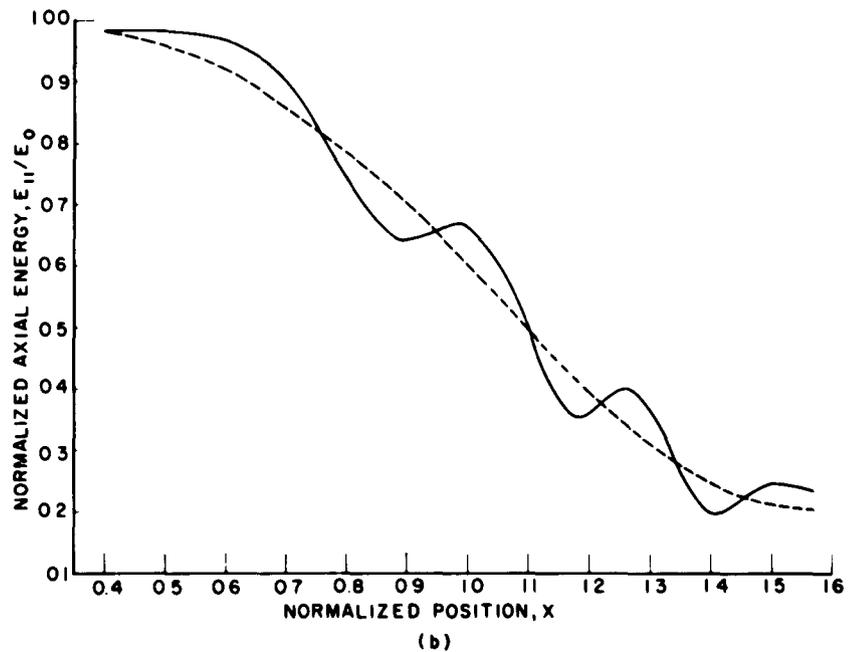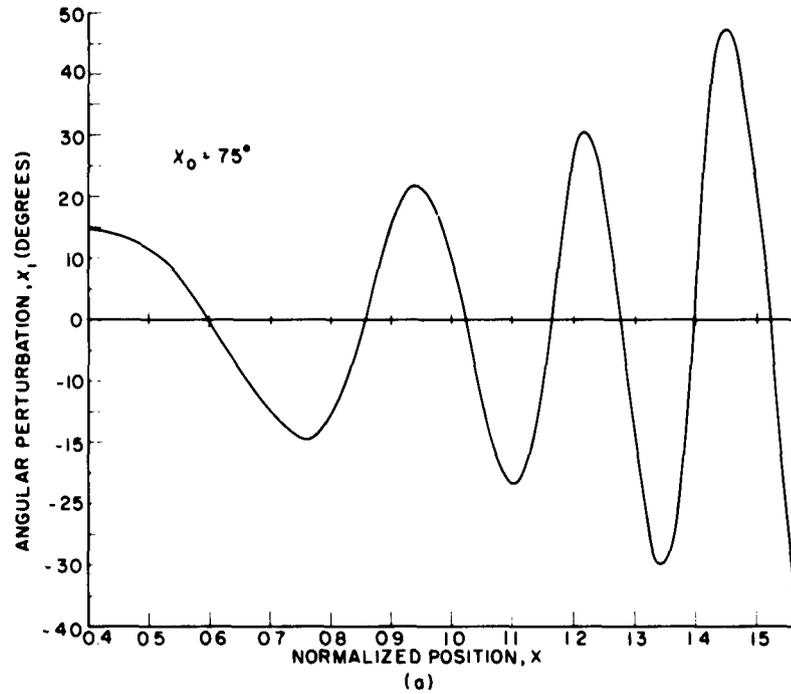
Fig. XII-25. (a) Angular deviation of particle position from the equilibrium orbit as a
function of the normalized distance, x. x = πz/2L, where z is the
position of the particle, and L is the length of the helical field.
(b) Normalized z-directed energy of the particle as a function of x. The
dashed line illustrates the behavior of an unperturbed particle.

Equations 14 and 15 were solved numerically by four-step Runge-Kutta integration. The results presented below were all computed by using a normalized distance increment of 10^{-3} (1570 steps for a single traverse). The results at the end of a complete pass were changed at most by 2 parts in the fifth significant figure for distance increments of 10^{-2} and 10^{-4}. A complete traverse, including computation and print-out of position, velocity, and energy at 157 points, requires approximately 21 seconds of IBM 7090 computer time.

4. Numerical Results

The equations of the system are characterized by three parameters:

α the ratio of exit azimuthal velocity to the total velocity of the particle;

γ a number proportional to the number or "corkscrew turns" in the system ($\gamma = 4L/P_0$); and

χ_0 the particle phase for which the corkscrew was designed.

Numerical computations were carried out for systems comparable to the electron corkscrew experiments of Wingerson and Dreicer ($\gamma = 20$), and for the more finely tuned systems of ultimate interest. We considered alpha values of 0.500 and 0.894 (25 per cent and 80 per cent of the energy in transverse motion at the exit), gamma values of 50 and 100, and design angles of 30°, 45°, 75°, and ±135°. The last two values are for the stable and unstable quadrants for an accelerating corkscrew ($\pi/2 \le x \le \pi$). The gamma values are those of an electron corkscrew with, for example, 200-gauss main axial field, 1600-volts injection energy, and 53-cm or 106-cm length. The number of turns depends on the value of α. For $\alpha = 0.894$ and $\gamma = 100$, this number is approximately 52. Some typical results are discussed below for this last case.

(a) Decelerating Corkscrews

A corkscrew as defined by Eq. 9 is decelerating for $0 \le x \le \pi/2$, and is phase stable for χ_0 in the first quadrant. Figure XII-25a shows the effect of a 15° perturbation in the angular position at $x = 0.4$. As expected, the oscillations grow in magnitude and frequency along the corkscrew. Figure XII-25b is a plot of axial energy during these oscillations compared with the smooth deceleration of an unperturbed particle. The angular-position perturbation leads to an energy perturbation because a particle in the first quadrant displaced forward in χ is subject to a smaller radial field, and is decelerated less strongly than an unperturbed particle.

Figure XII-26 illustrates the effect of an energy perturbation ($\Delta E = -0.06$) at the same position. The phase angle χ at various axial positions is marked off along the curve. The particle was lost when χ became more negative then $-45°$ at a time when
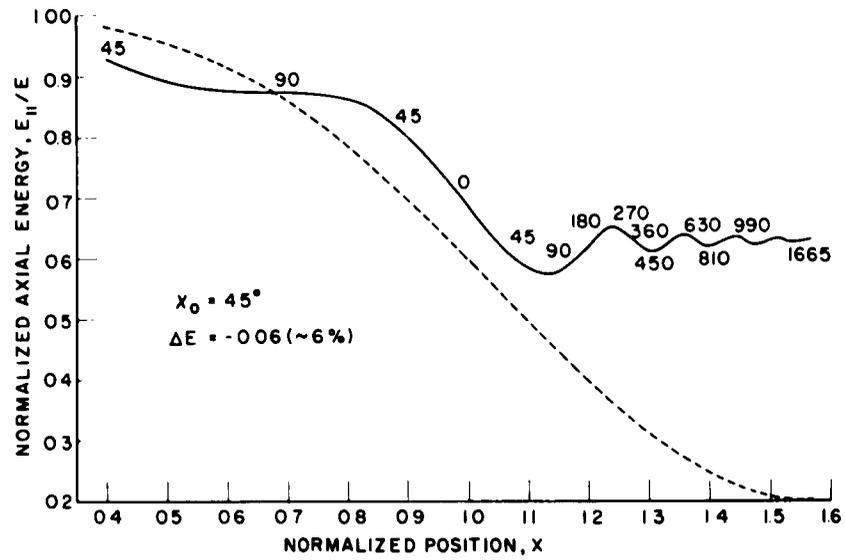
Fig. XII-26. Normalized z-directed energy versus x. The particle phase, χ, is plotted as a running parameter.
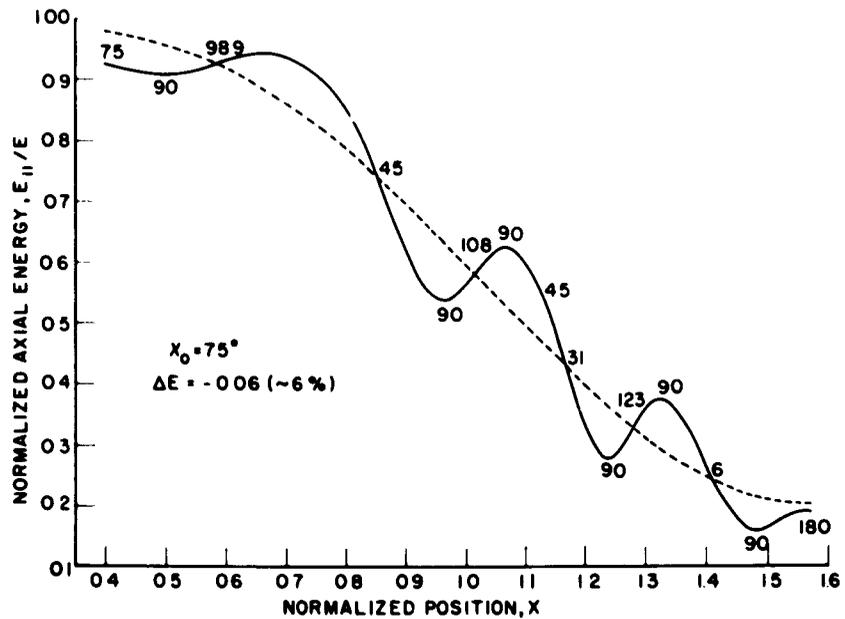
Fig. XII-27. Normalized z-directed energy versus x. The conditions for this case differ from those of Fig. XII-26 only in the value of χ_0. χ is the running parameter.

the axial velocity was higher than the axial velocity for an unperturbed particle. The radial field for $\chi < -45°$ is smaller than the design field, and the particle slipped even farther behind in phase. The radial field is negative for $-3\pi/2 \leqslant \chi \leqslant \pi/2$, and the particle was accelerated. It continued to slip behind in phase because $\frac{d\phi}{dz} = \frac{1}{v_z}\frac{d\phi}{dt}$ was everywhere smaller than $d\theta/dz$ as defined by the corkscrew windings. Successive cycles of acceleration $(-3\pi/2 + n\pi \leqslant \chi \leqslant \pi/2 + n\pi)$ and deceleration $(-\pi/2 \pm n\pi = \chi = \pi/2 \pm n\pi)$ followed, but the effect of each became smaller because the particle spent less time in the successively shorter sections.

Figure XII-27 is a plot of the same case for an initial design phase, χ_o, of $75°$. Two factors cause this situation to be more stable than that in Fig. XII-26. First, the particle must slip farther back in phase to see a radial field weaker than the design field. Second, and more important, the ratio of maximum radial field to design field is larger for $\chi_o = 75°$ than it is for $\chi_o = 45°$. $[\omega_r(\chi = 0) = 1.4 \omega_r(45°) = 3.9 \omega_r(75°)]$. The improved stability is, of course, accompanied by increased perturbation of previously trapped particles.

(b) Accelerating Corkscrews

The pitch equation (9) describes an accelerating corkscrew for $\pi/2 \leqslant x \leqslant \pi$. A system with parameters identical to the decelerating corkscrew discussed above ($a = 0.894$, $\gamma = 100$) was investigated for the case $\chi_o = 135°$ and $x > \pi/2$. The effect of an angular perturbation at the entrance is depicted in Fig. XII-28. The result of an identical
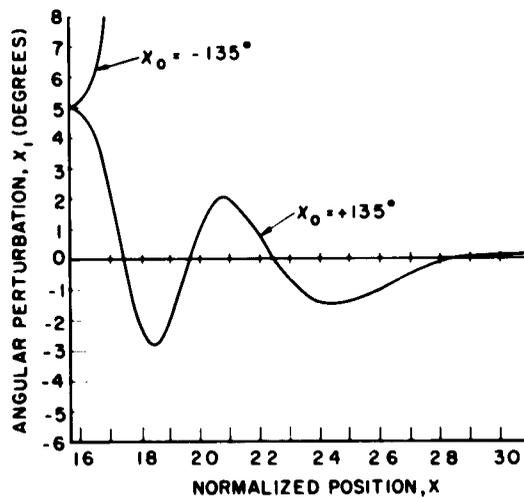
Fig. XII-28. Angular deviation versus x for an accelerating corkscrew. Note that $\pi/2 \leqslant x \leqslant \pi$ for this case.
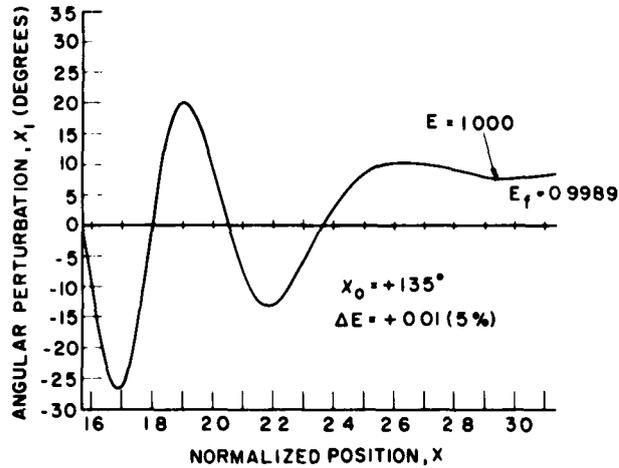
Fig. XII-29.   Angular deviation versus x for the corkscrew of Fig. XII-28.
In this case, the initial perturbation was in energy.

perturbation for a particle in the unstable accelerating quadrant $(x_0 = -135°)$ is shown. As predicted, the oscillations of the particle in the proper quadrant are rapidly damped. The angular oscillations resulting from a 5 per cent energy perturbation at the entrance are shown in Fig. XII-29. The particle was completely unwound at the local minimum at $x = 2.95$. The longitudinal energy at the exit was 0.9989 of the total energy.

5. Conclusions

The numerical results presented for the decelerating corkscrew thread the passage between the rock of experiment — the corkscrew decelerates particles — and the whirlpools of intuition and linearized theory — the decelerating corkscrew is unstable to growing oscillations. The corkscrew is indeed unstable, but not too unstable. The numerical results agree nicely with the experimental work of Dreicer and his co-workers[2] in showing that a small energy spread at the exit is accompanied by a large spread in exit phase. The size of the "acceptance hole" at the corkscrew entrance is also in agreement with experiment ($\pm$ ~10 per cent in energy, $\pm$ ~30° in phase).

Cases similar to that shown in Fig. XII-26, which indicates that particles can be partially wound up to an extent determined by their initial deviation from the stable orbit, call into question an earlier conclusion that the optimum system would consist of a finely tuned corkscrew with a very carefully matched injection system. It is likely that an injection system with a moderate spread in particle velocity (resulting in a wide spread in energy and phase at the corkscrew exit) would circumvent many of the difficulties that plague devices with highly ordered particle motions.

The accelerating corkscrew presents intriguing possibilites for injection into

"closed" toroidal systems. In the simple form discussed in this report it is capable of placing particles on the axis of a toroidal system whose minor radius is equal to the cyclotron radius of a particle with total velocity in the transverse direction. Injection into a field extending over several cyclotron radii is more desirable. Modifications of the accelerating corkscrew to accomplish this end are under study.

L. M. Lidsky

References

1. R. C. Wingerson, Phys. Rev. Letters 6, 446 (1961); R. C. Wingerson, Quarterly Progress Report No. 60, Research Laboratory of Electronics, M. I. T., January 15, 1961, pp. 59-65.

2. H. Dreicer et al., Nuc. Fus. 1962 Supplement I, p. 299, 1962.

3. R. C. Wingerson, T. H. Dupree, and D. J. Rose (unpublished material).

G. FUSION REACTION BLANKET EXPERIMENT: ANALYSIS OF THRESHOLD-DETECTOR DATA

The neutron-energy spectra in mock-ups of a tritium regenerating blanket to surround a power-producing deuterium-tritium-cycle thermonuclear reactor will be measured by means of the following threshold reactions: $U^{238}(n,f)$, $P^{31}(n,p)Si^{31}$, $Fe^{56}(n,p)Mn^{56}$, $I^{127}(n,2n)I^{126}$, and $F^{19}(n,2n)F^{18}$. A description of the experimental arrangement and of the physical properties of the threshold-detector foils has been given previously.[1] The method of analysis of the activities from these reactions for the purpose of determining the neutron spectrum, generalized to the case of N different threshold reactions, is given in this report.

The problem is the solution of N simultaneous integral equations for the neutron spectrum $\phi(E)$:

$$T_j = \int \phi(E)\sigma_j(E)\, dE, \qquad (j=1,2,\ldots,N). \tag{1}$$

Here, the T_j is the activation rate per atom of the jth threshold detector corrected for decay, and $\sigma_j(E)$ is the activation cross section. The range of all integrals in this discussion is from the lowest threshold energy to the maximum energy to be expected in the experiment. Since the general solution of Eq. 1 is intractable, some assumptions about the $\sigma_j(E)$ and the neutron-energy spectrum $\phi(E)$ must be made. The method of Lanning and Brown[2] has been chosen for two reasons: it is relatively simple, and it allows additional knowledge of the spectrum — in this case, from calculations made with machine codes developed by Impink[3] — to be taken into account by means of a weight function $w(E)$.

In this method $\phi(E)$ is expanded as a product of $w(E)$ times a weighted sum of polynomials $p_i(E)$:

$$\phi(E) \sim w(E) \sum_{i=1}^{N} a_i p_i(E). \tag{2}$$

The polynomials $p_i(E)$ (of degree i-1) are chosen so that

$$\int p_i(E) p_k(E) w(E) \, dE = \delta_{ik}. \tag{3}$$

The $\sigma_j(E)$ are approximated as a Fourier expansion of the $p_i(E)$:

$$\sigma_j(E) \sim \bar{\sigma}_j(E) = \sum_{i=1}^{N} \tau_{ji} p_i(E), \tag{4}$$

where

$$\tau_{ji} = \int p_i(E) \sigma_j(E) w(E) \, dE. \tag{5}$$

The method gives a least-squares fit to the cross-section data:

$$\int w(E) \left[\sigma_j(E) - \bar{\sigma}_j(E) \right]^2 \, dE = \text{minimum}. \tag{6}$$

If we use these equations, the activation of the jth foil is given by

$$T_j = \sum_{j=1}^{N} \tau_{ji} a_i, \qquad (j=1, \dots, N). \tag{7}$$

The N equations (7) are solved for the a_i; this operation completes the calculation of all of the parameters that are necessary to determine $\phi(E)$ by Eq. 2.

A Fortran-II program has been written to perform the calculations, and has been tested by using activities computed from spectra calculated by Impink. Two examples are shown in Fig. XII-30 in which the input spectrum is shown with two spectra calculated from Eq. 2 by using two different $w(E)$ for two sets of data. Since only 5 foils are used, the calculated spectrum is dependent on the function chosen for $w(E)$. If the weight function $w(E)$ is chosen to be the input spectrum, of course, the agreement is excellent. As an example of a weight function that has the general shape of the input spectrum, but does not depend directly on the spectrum, a weight function of the form

$$w(E) = (A + BE^J)^{-1} + C \exp(-(E-F)^2/G^2) \tag{8}$$
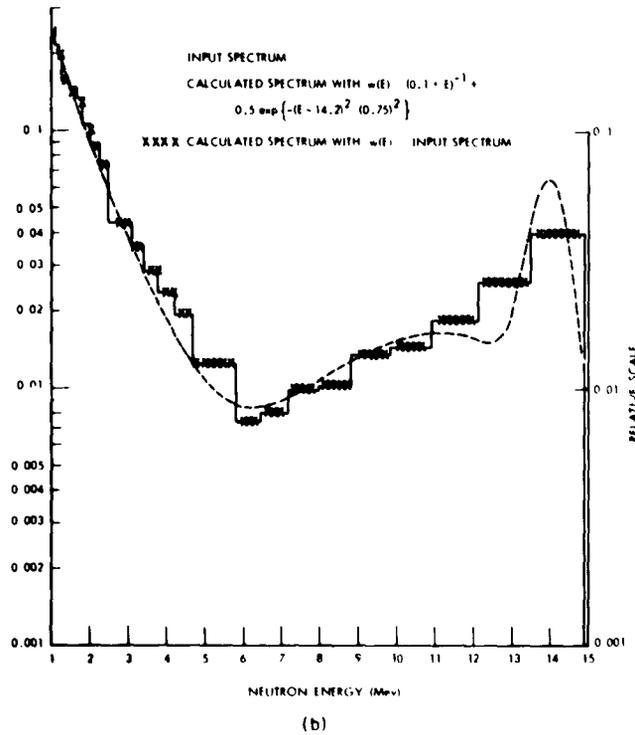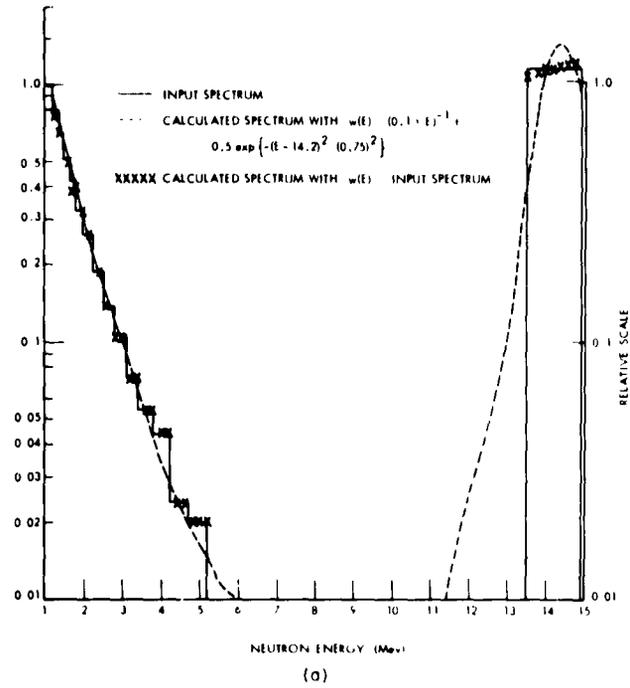
Fig. XII-30. Comparison of spectrum calculated from Eq. 2 with input spectrum calculated from Impink's codes. Data in (a) are from the first wall of the blanket; data in (b) are from a point approximately 20 cm from the first wall.

Table XII-1. Comparison of threshold activities.

Reaction	Threshold energy (Mev)	Used to determine a_i in Eq. 2	Calculated activity[a]	
			From input spectrum	From $\phi(E)$ according to Eq. 2
$U^{238}(n, f)$	1.1	yes	2264.1	2264.1
$P^{31}(n, p)Si^{31}$	1.6	yes	163.65	163.74
$Fe^{56}(n, p)Mn^{56}$	4.5	yes	226.95	226.95
$I^{127}(n, 2n)I^{126}$	9.5	yes	2103.0	2103.0
$F^{19}(n, 2n)F^{18}$	11.1	yes	87.808	87.808
$Al^{27}(n, \alpha)Na^{24}$	6.1	no	208.00	208.75
$Zn^{64}(n, p)Cu^{64}$	1.9	no	421.75	425.69
$Ni^{58}(n, 2n)Ni^{57}$	12.3	no	43.139	42.609

[a]Activity calculated by $T_i = \int \phi(E)\sigma_i(E) \, dE$, where $\sigma_i(E)$ is the cross section, and $\phi(E)$ is the input spectrum, or the spectrum calculated with w(E) from Eq. 8, by using data from Fig. XII-30a.

was chosen. The values of the constants A, B, C, F, and G were chosen somewhat arbitrarily; the fit of the calculated spectrum to the input spectrum is not too dependent upon the choice. Furthermore, the same $w(E)$ gives a good fit to the data at the first wall (Fig. XII-30a), where the peak centered about 14.2 Mev predominates, as well as farther in the blanket (Fig. XII-30b), where this peak has been attenuated, and the intervening energy regions have been filled in by the moderated neutrons.

Tests were also made with $w(E) = 1.0$. The results showed that the input spectrum is too complicated to be fitted by a fourth-order polynomial: The calculated spectrum, while it did show the trends at the lower and upper regions of the spectrum, oscillated about the origin in the region 6-12 Mev.

In these calculations the assumption was made that the T_j are known exactly, which is not actually true. In a series of test runs, the T_j calculated from the input spectra were varied by fixed amounts. The results showed that the shape of the spectrum calculated from Eq. 2 is most sensitive to the activities of the reactions with intermediate threshold energies: $P^{31}(n,p)$, $Fe^{56}(n,p)$, and especially $I^{127}(n,2n)$. Increasing or decreasing the activities of the foils with the highest and lowest threshold energies merely increases or decreases the value of the calculated spectrum in the corresponding energy range.

Since these calculations are based on an integral method, information about $\phi(E)$ in regions where it is small (say <0.01 of its maximum value) can be lost in the uncertainties in the method; likewise for $\sigma_i(E)$ in regions where $\phi(E)$ is large. Thus a fairer test of the method is its ability to reproduce activation rates of reactions which were not used [1] in determining the calculated spectrum because in performing the integral $\int \phi(E)\sigma(E)\, dE$ to determine these activities, errors in regions where $\phi(E)$ is small are of less importance. The activities of three such reactions, obtained by integrating the cross section for the reaction with the input and calculated spectra are shown in Table XII-1. The maximum deviation between the activities obtained from the input spectrum, and those obtained from the spectrum calculated with the Gaussian weight function, is approximately 1 per cent.

P. S. Spangler

References

1. P. S. Spangler, Fusion reactor blanket experiment. Quarterly Progress Report No. 69, Research Laboratory of Electronics, M. I. T., April 15, 1963, pp. 81-88.

2. W. D. Lanning and K. W. Brown, Report WAPD-T-138, Bettis Atomic Power Laboratory, Pittsburgh, Pennsylvania, 1961; see also Trans. Am. Nuclear Soc. 4, 267 (1961).

3. A. J. Impink, Neutron Economy in Fusion Reactor Blanket Assemblies, Sc.D. Thesis, Department of Nuclear Engineering, M. I. T., January 1963.

H. EMISSION OF CYCLOTRON HARMONICS FROM PLASMAS

Considerable interest has been stimulated recently by the observation of maxima in the emission and absorption spectrum of plasma at harmonics of the cyclotron frequency.[1] Since the electron energies in the plasmas were nonrelativistic, some mechanism other than pure single-particle cyclotron radiation must be responsible for the observed spectrum. Simon and Rosenbluth[2] have suggested that the radiation is due to electrons whose trajectories pass through sheaths. The electric fields in the sheaths distort the circular Larmor orbits, and the result is the emission of radiation at the cyclo' on harmonics.

The qualitative aspects of the observed spectrum can also be produced by another mechanism: the scattering of longitudinal electron waves by the ions. At equilibrium, such scattering produces a relatively smooth bremsstrahlung spectrum, but for a non-equilibrium plasma, the spectrum can be quite different. In particular, for a plasma in a magnetic field, a longitudinal wave traveling perpendicular to the magnetic field can produce scattered radiation at the cyclotron harmonics.

Given the spectrum of electron and ion fluctuations, the calculation of the scattering is straightforward but too long to record here. Only the results will be given. We shall take the externally imposed magnetic field to lie along the z axis with an electron cyclotron frequency ω_c. We shall assume that the spectrum of (random) longitudinal electron waves (fluctuations) has a magnitude that generally is greater than the equilibrium level. The ions are assumed to have infinite mass, but to be correlated with each other. The average electron and ion densities are n_e and n_i, respectively. The fluctuations of

electron and ion density are denoted $n_e \delta \rho_e(\vec{r}, t)$ and $n_i \delta \rho_i(\vec{r})$. The spectral density of electron fluctuations, $\langle \delta \rho_e \delta \rho_e | \vec{K}, \omega \rangle$, is defined as

$$\langle \delta \rho_e \delta \rho_e | \vec{K}, \omega \rangle = \int_{-\infty}^{\infty} d\tau \, e^{-i\omega\tau} \int d\vec{a} \, e^{-i\vec{K}\cdot\vec{a}} \langle \delta \rho_e(\vec{r}, t) \, \delta \rho_e(\vec{r}+\vec{a}, t+\tau) \rangle.$$

Here, the angular brackets denote an ensemble average. In similar fashion, we define the spectral density of ion fluctuations as

$$\langle \delta \rho_i \delta \rho_i | \vec{K} \rangle = \int d\vec{a} \, e^{-i\vec{K}\cdot\vec{a}} \langle \delta \rho_i(\vec{r}) \, \delta \rho_i(\vec{r}+\vec{a}) \rangle.$$

The electron fluctuations will be scattered by the ion fluctuations, and produce longitudinal and transverse current-density fluctuations whose spectral density $\langle \delta \vec{j} \, \delta \vec{j} | \vec{k}, \omega \rangle$ is given (approximately) by

$$\langle \delta \vec{j}_r \delta \vec{j}_s | \vec{k}, \omega \rangle = \int d\vec{K} \, g_{rs} |b(\vec{K}, \omega)|^2 \langle \delta \rho_e \delta \rho_e | \vec{K}, \omega \rangle \langle \delta \rho_i \delta \rho_i | -\vec{K} \rangle,$$

where

$$g_{xx} = K_\perp^2 \frac{\omega^2 + \omega_c^2}{\left(\omega^2 - \omega_c^2\right)^2}$$

$$g_{zz} = K_z^2 \frac{1}{\omega^2}$$

$$g_{xz} = 0$$

$$b(\vec{K}, \omega) = K^{-4} \sum_{m=-\infty}^{+\infty} m\omega_c \omega_p^4 n_e \int d\vec{v} \, \frac{J_m^2\left(\frac{K_\perp v_\perp}{\omega_c}\right) \frac{\partial}{\partial v^2} f_e(v^2)}{\omega + K_z v_z + m\omega_c - i\epsilon}.$$

Here, ω_p is the electron plasma frequency, J_m is a Bessel function, and $K_\perp^2 = K_x^2 + K_y^2$.

The emission power density of scattered radiation into the mode $(\vec{k}, \omega)$ is proportional to $\langle \delta \vec{j} \delta \vec{j} | \vec{k}, \omega \rangle$. The actual flux reaching a receiver outside the plasma depends, of course, on the dielectric properties of the plasma, including the reflection at the boundaries. One may expect (at least for an optically thin plasma) that the gross features of the spectrum of $\langle \delta \vec{j} \delta \vec{j} | \vec{k}, \omega \rangle$ will be present in the observed spectrum.

For convenience, we shall take $\langle \delta \rho_i \delta \rho_i | \vec{K} \rangle$ to be the familiar equilibrium expression

$$n_i^{-1} \frac{K^2 + \frac{1}{2}\lambda_D^{-2}}{K^2 + \lambda_D^{-2}}.$$

If $K_z \ll \lambda_D^{-1}$ and $K_\perp \approx \lambda_D^{-1}$, then $\left|b(\vec{K},\omega)\right|^2$ has maxima at $\omega = m\omega_c$. Now suppose that when $K_\perp \approx \lambda_D^{-1}$, $\left\langle \delta\rho_e \delta\rho_e \middle| \vec{K},\omega \right\rangle$ is large only for $K_z \ll \lambda_D^{-1}$. Then $\left\langle \delta\vec{j}\delta\vec{j} \middle| \vec{k},\omega \right\rangle$ will also have the observed maxima at the cyclotron harmonics. Furthermore, this choice of the electron fluctuation spectrum is not entirely artificial. One might expect just such a spectrum in situations in which the unstable modes are those that travel in a direction approximately perpendicular to the magnetic field so that the Landau damping is at a minimum.

Since $\left\langle \delta\rho_e \delta\rho_e \middle| \vec{K},\omega \right\rangle$ should not be large for $\omega > \omega_p$, one would not expect to observe harmonics at frequency greater than ω_p. This prediction is in agreement with experiments.

<div style="text-align:right">T. H. Dupree</div>

References

1. G. Bekefi, J. D. Coccoli, E. B. Hooper, Jr., and S. J. Buchsbaum, Phys. Rev. Letters 9, 6 (1962).

2. A. Simon and M. N. Rosenbluth (paper to be published in Phys. Fluids).

I. ERRATA: SCATTERING OF LIGHT FROM (PLASMA) ELECTRONS II

In a report with this title, published in Quarterly Progress Report No. 69 (pages 74-79), corrections should be made in two equations.

Page 76 — the third and fourth equations from the top of the page should read:

$$V^* = \frac{\Delta\lambda_2}{\lambda_o} c \frac{1}{1 + \cos\theta}.$$

Hence

$$dN_{V_a} = \frac{N}{\sqrt{2\pi}} \left(\frac{m}{2kT_e}\right)^{1/2} \frac{c}{\lambda_o} (1+\cos\theta)^{-1/2} \exp\left(-\frac{mc^2(1+\cos\theta)^{-1}}{4k\lambda_o^2} \frac{\Delta\lambda_2^2}{T_e}\right) d\lambda,$$

<div style="text-align:right">E. Thompson, G. Fiocco</div>

J. CONCENTRATION OF EXCITED STATES IN A LOW-ENERGY CESIUM PLASMA

1. Introduction

In a low-energy cesium plasma ions are lost continuously by ambipolar diffusion and recombination. In the absence of a sufficient external ion source these losses have to be balanced by ionization processes occurring in the plasma. Ionization occurs mainly by inelastic collisions between electrons and cesium atoms. With electrons of low energy most of these collisions lead to the first excited state of the cesium atom, that is, the normal 6s level of the valence electron is excited to the 6p level. Ionization can then occur by further collisions of these excited atoms. In this report the concentration of atoms in the first excited state is calculated as a first step toward obtaining the rates of ionization from excited states.

The concentration of excited states is mainly determined by excitation and de-excitation collisions with electrons and by emission and absorption of resonance radiation. The results of calculations on the required electron collision cross sections are given in this report. The problem of radiation trapping, that is, the successive emission and absorption of resonance radiation, is also treated. An over-all rate balance is used to determine the concentration of excited states.

The calculations are carried out for a cesium density of 2×10^{16} cm^{-3}, a cesium temperature of $1000\,°K$, an electron density of 10^{14} cm^{-3}, electron temperature of approximately $3000\,°K$, and a plasma thickness of 0.025 cm. These values are typical for the high-current mode of a cesium thermionic-energy converter and can be varied considerably without affecting the conclusions.

2. Electron Cross Sections

The cross-section curve for excitation of cesium atoms to the first excited state by electron collisions has not been measured. The excitation curve has been measured, however, and calculated theoretically for sodium, which is very similar to cesium in its atomic properties. Because of this similarity, the same theoretical method that yielded good agreement with the experimental data for sodium is used to calculate the inelastic cross-section curve for cesium.

The total electron collision cross section represents an upper limit to the inelastic cross section. Total cross-section curves were measured by Brode[1] for the alkali metals — sodium, potassium, rubidium, and cesium. All of the curves have similar shapes, showing a peak at electron energies of approximately 2 ev and decreasing at higher energies. The curve for cesium shows the highest cross sections, with a peak value of 570×10^{-16} cm^2. This leads to the expectation that the inelastic electron cross sections for cesium will also be large and will be similar to, but somewhat higher than,

those for the other alkali metals.

The general shape of the electron cross-section curve for excitation from the ground state is well known. The cross section is zero up to the excitation potential; then it rises sharply with increasing electron energy, levels off to a peak a few volts above the excitation potential, and then declines. Such curves have been measured for sodium, helium, and mercury.[2]

In this report we are interested mainly in the initial rise of the excitation curve because there are very few electrons in the plasma with energies higher than a few ev. In the low-energy region the normal Born approximation yields cross sections that are too large. Recently, Seaton[3] has developed a semi-classical "impact-parameter" method for calculating electron excitation cross sections of optically allowed transitions. This method yields results that are in good agreement with a variety of available experimental data at low electron energies. To see this point clearly, let us consider the case of the excitation of sodium to the first excited state, that is, the 3s-3p transition.

In the impact-parameter method the motion of the colliding electron is taken to be rectilinear. The cross section is calculated in terms of the contributions from different impact parameters R, where R is the classical distance of closest approach. With initial state i and impact parameter R_i, let there be a probability $P_{ji}(R_i)$ that the i → j transition occurs. The cross section is

$$\sigma(i \to j) = \int_0^\infty P_{ji}(R_i)\, 2\pi R_i\, dR_i. \tag{1}$$

The probability of transition P_{ji} is calculated from time-dependent perturbation theory, under the assumption that the atomic potential follows the simple Coulomb law. This assumption is valid for $R_i \gg r_a$, where r_a is a length that is comparable with the atomic dimensions, but it is invalid for $R_i \lesssim r_a$. For some transitions with large cross sections such as the 3s-3p transition in sodium the calculated values of P_{ji} violate the conservation condition, $P_{ji} \lesssim 1$, for the smaller impact parameters. For these transitions the impact-parameter method introduces a cutoff in Eq. 1 at an impact parameter R_1 which is such that

$$P_{ji}(R_1) = 1/2. \tag{2}$$

For $R_i < R_1$ the probability P_{ji} is considered as an oscillatory function with a mean value of 1/2. The cross section then is

$$\sigma(i \to j) = \frac{1}{2}\pi R_1^2 + \int_{R_1}^\infty P_{ij}(R_i)\, 2\pi R_i\, dR_i. \tag{3}$$

Seaton[3] used the impact-parameter method to calculate the inelastic scattering cross

section for excitation to the first excited state of the sodium atom, that is, for the 3s-3p transition. These calculations were repeated and resulted in the cross-section curve shown in Fig. XII-31. This figure also shows absolute cross-section measurements by Christoph[4] as corrected by Bates,[2] and relative cross-section measurements by Haft[5] as reported by Christoph,[4] scaled to agree with the absolute measurements. Haft's data as reported here differ somewhat from the curve drawn by Bates,[2] who does not seem to consider Haft's low-energy points and scaled the data to agree with Christoph's absolute measurements at high electron energies, for which Haft's values are too high because of Doppler broadening of the spectral lines that he observed.

As shown in Fig. XII-31, the impact-parameter approximation gives very good agreement with the experimental data down to 0.8 ev above the excitation potential, which is 2.1 ev for sodium. At these very low energies, Haft's data are not very accurate, but the agreement is actually better than shown, since in Haft's experiment the electron energies had a spread of approximately 0.4 ev below their nominal value. At the excitation potential, the impact approximation curve is clearly in error, since it does not go to zero there. The Born approximation curve is much too high at low energies.

The good agreement between theory and experiment for the 3s-3p excitation of sodium encouraged the application of the theory to a very similar case, the 6s-6p excitation of cesium. The two 6p levels of cesium lie close together approximately 1.4 ev above the ground state. As for sodium, the two levels were treated together and yielded the excitation cross-section curve shown in Fig. XII-32. The curve is very similar to that for sodium; the maximum is higher and occurs at a lower energy, as expected in view of the
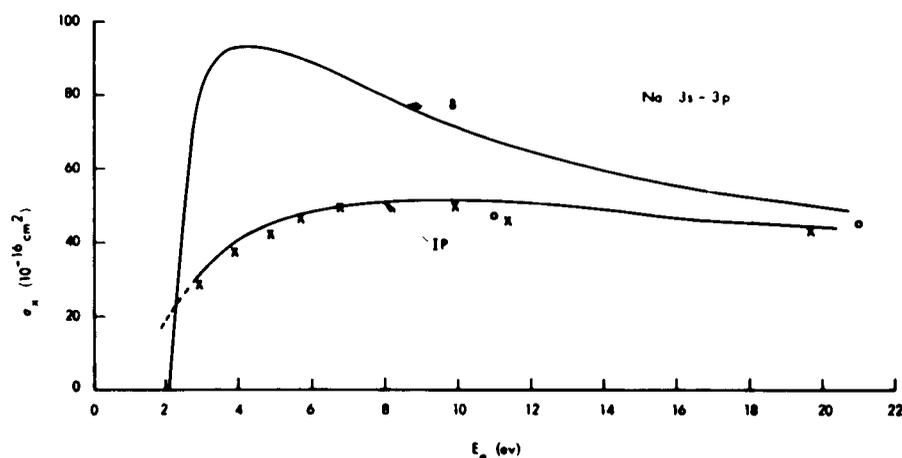
Fig. XII-31. The 3s-3p inelastic electron collision cross section of sodium as a function of electron energy. B, Born approximation; IP, impact-parameter approximation; O, experimental (Christoph, absolute); x, experimental (Haft, relative).

larger total collision cross section and lower excitation energy of cesium. Judged from the data for sodium, the curve for cesium should be correct down to electron energies of 2.2 ev. For lower electron energies a conservatively low estimate of the dependence of the excitation cross section on electron energy E is

$$\sigma_x(E) = 75 \times 10^{-16}(E-1.4) \ cm^2$$

(4)

$$1.4 \leqslant E \leqslant 2.2 \ ev.$$

Once the excitation cross section is known, the de-excitation cross section can be calculated by invoking the principle of detailed balance. According to this principle, the rates of electron excitation and de-excitation must be equal under thermal equilibrium at all temperatures. It has been shown directly by Fowler[6] that the relation between the de-excitation cross section σ_d and the excitation cross section σ_x is given by

$$\frac{\sigma_d(E)}{\sigma_x(E+E_x)} = \frac{\omega_o}{\omega_x} \frac{E + E_x}{E},$$

(5)

where E_x is the excitation potential and ω_o and ω_x are the statistical weights of the ground state and the excited state. The de-excitation cross section for cesium 6p-6s has been calculated and is shown in Fig. XII-32. It is approximately constant in the energy range plotted.
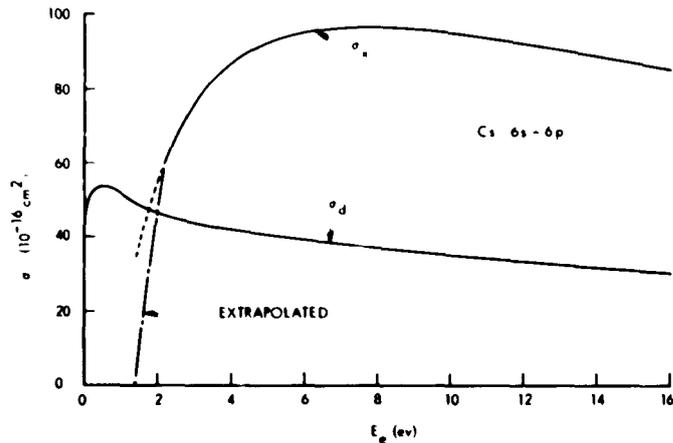
Fig. XII-32. The 6s-6p excitation and de-excitation electron collision cross section of cesium as a function of electron energy. Impact-parameter approximation.

3. Radiation Trapping

One source of losses of excited atoms in a cesium plasma is the spontaneous emission of resonance radiation. This radiation consists of sharp lines that have a high probability of absorption by atoms in the ground state. The process of successive absorption and emission of resonance radiation, called "radiation trapping," results in an effective lifetime that is larger than the natural lifetime of the excited state.

The total rate of decay of excited states by spontaneous emission is n_x/τ_n cm^{-3} sec^{-1}, where n_x is the concentration, and τ_n the natural lifetime of the excited state. Suppose that the emitted radiation has an average probability p of escaping from the plasma without absorption. Then the rate at which these photons leave the plasma is

$$\nu_r = \frac{n_x p}{\tau_n} \text{ cm}^{-3} \text{ sec}^{-1}. \tag{6}$$

This is equal to the net radiation loss of excited states per unit volume. Equation 6 can be written

$$\nu_r = n_x/\tau_{eff}, \qquad \tau_{eff} = \tau_n/p, \tag{7}$$

where τ_{eff} is the effective lifetime for radiative decay of the excited state, averaged over the plasma. If the photon-escape probability p is small, then τ_{eff} will be much larger than τ_n. In this report we are concerned only with low radiation densities, where radiation-stimulated emission is not important.

An expression for the average radiation-escape probability will be derived from the absorption cross section and the emission line shape, for a cesium plasma between two semi-infinite flat plates. Any radiation reaching the plates will be considered to be lost. It will be seen that the calculations can be easily extended to other resonance lines if certain physical data are available.

a. Photon Absorption Cross Section and Emission Probability

In calculating the average radiation-escape probability p we must consider the detailed shape of the emitted line and of the corresponding absorption cross section. Several disturbing influences cause these shapes to differ from a pure delta function; these influences are said to broaden the lines. Under thermodynamic equilibrium the shapes of the absorption and emission lines are the same, since detailed balancing between emission and absorption must hold at all frequencies. This means that the effect of a broadening mechanism must only be known well for either absorption or emission in order to calculate the photon-escape probability.

In the plasma under consideration the dominant broadening mechanism is pressure broadening. It is due to collisions between excited states and ground-level atoms of the

same species that cause broadening, frequency shift, and asymmetry of the absorption cross section.[7] The escape probability of photons is determined by the line shape in the wings of the line, far from the center. In this region the absorption shape follows closely the simple Breit-Wigner formula:

$$\sigma_\nu = \sigma_o \frac{(\Gamma_p/2)^2}{(\Gamma_p/2)^2 + (\nu-\nu_o)^2}. \tag{8}$$

Here, ν is the radiation frequency, σ_o is the absorption cross section at the resonance frequency ν_o, and Γ_p is the halfwidth (full width at half-maximum) resulting from pressure broadening. The absorption cross section at the edges of the resonance curve of the cesium 6s-6p transition has been measured by Gregory.[8] His data agree well with the theoretical result of Furssow and Wlassow,[9] who calculated the halfwidth resulting from pressure broadening to be

$$\Gamma_p = \frac{2}{3\pi} \frac{e^2}{m_e \nu_o} n_o f, \tag{9}$$

where n_o is the density of atoms in the ground state and f is the oscillator strength for the transition.

The value of σ_o can be obtained from a consideration of thermodynamic equilibrium between the radiation and the atoms. The derivation has been given by several authors.[10,11] It can be shown that the integral of the absorption cross section over the resonance is proportional to the oscillator strength, f:

$$\int \sigma_\nu \, d\nu = \frac{\pi e^2}{m_e c} f. \tag{10}$$

Combining Eqs. 8-10 yields

$$\sigma_\nu = \frac{3\pi}{n_o \lambda_o} \frac{1}{1 + t^2}, \tag{11}$$

where λ_o is the wavelength at the resonance frequency ν_o and

$$t = \frac{\nu - \nu_o}{\Gamma_p/2}. \tag{12}$$

The emission probability $p(\nu)$ is just the Breit-Wigner shape scaled to give a total emission probability of unity:

$$p(\nu) = \frac{1}{\pi} \frac{\Gamma_p/2}{(\Gamma_p/2)^2 + (\nu-\nu_o)^2}. \tag{13}$$

b. Calculation of the Photon Escape Probability

Consider a semi-infinite plasma of dimension s, as shown in Fig. XII-33.

Suppose that an excited atom at position x decays by emitting a photon of frequency ν in the direction r. The excape probability for this photon is $\exp(-n_0 \sigma_\nu r)$. To obtain the average excape probability over the emission spectrum one must multiply this exponential
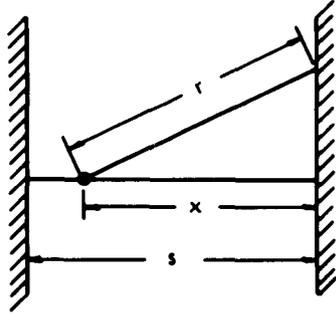
Fig. XII-33. Schematic diagram of the plasma geometry.

factor by the emission probability $p(\nu)$ and integrate over all frequencies. Similarly, the average over all directions of emission is obtained by multiplying by $p(r)$, the probability of emission in a direction between r and r + dr, and by integrating over all r. Finally, this is averaged over all values of x. The average escape probability then is given by

$$p = \frac{1}{s} \int_0^s dx \int_x^\infty p(r)\, dr \int_0^\infty p(\nu)\, e^{-n_0 \sigma_\nu r}\, d\nu. \tag{14}$$

The photons are emitted with equal probability in all directions. The directional emission probability is

$$p(r) = \frac{x}{r^2} \qquad \int_x^\infty p(r)\, dr = 1. \tag{15}$$

Combining Eqs. 11-15 yields

$$p = \frac{1}{\pi s} \int_0^s x\, dx \int_x^\infty \frac{dr}{r^2} \int_{-\infty}^\infty \frac{dt}{1+t^2} \exp\left(-3\pi \frac{r}{\lambda_0} \frac{1}{1+t^2}\right). \tag{16}$$

This integral is rather difficult to evaluate. A first approximation can easily be obtained, since only photons with frequencies far out in the wings contribute significantly to the average escape probability. For these photons $t^2 \gg 1$. If we neglect 1 in comparison with t^2 in Eq. 16 the integral is readily evaluated to be

$$p = \frac{4}{3\pi}\sqrt{\frac{\lambda_o}{3s}}. \tag{17}$$

A second approximation to the integral in Eq. 16 was obtained by separating it into two integrals: one for low values of r, and one for high values. The result is

$$p = \frac{4}{3\pi}\sqrt{\frac{\lambda_o}{3s}}\left(1 - \frac{1}{11}\sqrt{\frac{\lambda_o}{s}}\right). \tag{18}$$

Since $\lambda_o \ll s$, the correction introduced is negligible and the first approximation, Eq. 17, gives the average escape probability with sufficient accuracy. This result shows that we were justified in ignoring the detailed shape of the line near the center, since this region contributes very little to the escape probability.

c. Calculation of the Effective Lifetime of the Excited States

Equations 7 and 17 give the effective lifetime as

$$\tau_{eff} = \tau_n \frac{3\pi}{4}\sqrt{\frac{3s}{\lambda_o}}. \tag{19}$$

For the cesium 6p state, $\tau_n = 3.5 \times 10^{-8}$ sec, and $\lambda_o = 0.87 \times 10^{-4}$ cm. With a plasma spacing, s, of 0.025 cm, the effective lifetime is

$$\tau_{eff} = 2.4 \times 10^{-6} \text{ sec.} \tag{20}$$

It is interesting to note that the effective lifetime is independent of the cesium atom density. This is only true as long as cesium atoms are the dominant source of resonance line broadening.

4. Rate Balances

The concentration of excited states is primarily determined by the rate of excitation v_x ($cm^{-3} sec^{-1}$), the rate of de-excitation v_d, and the net rate of radiative decay v_r. The last rate is determined by the effective lifetime τ_{eff} of the excited states. The rate balance is

$$v_x = v_d + v_r. \tag{21}$$

The rate of de-excitation is

$$v_d = n_x n_e \int_0^\infty f_e(E) \, \sigma_d(E) \, v_e(E) \, dE, \tag{22}$$

where $f_e(E)$ is the electron energy distribution function, and n_e and v_e are the electron

density and velocity. Since σ_d is approximately constant in the low-energy range of interest, Eq. 22 can be written

$$\nu_d = n_x n_e \bar{\sigma}_d \bar{v}_e.$$

Combining this with Eqs. 7 and 21 yields

$$\frac{\nu_x - \nu_d}{\nu_d} = \frac{1}{\tau_{eff} n_e \bar{\sigma}_d \bar{v}_e} \equiv R, \tag{23}$$

where $\bar{\sigma}_d = 50 \times 10^{-16}$ cm^2, $\bar{v}_e = 3 \times 10^7$ cm/sec, $\tau_{eff} = 2.4 \times 10^{-6}$ sec, and $n_e = 10^{14}$ cm^{-3}. With these values, Eq. 23 becomes

$$\frac{\nu_x - \nu_d}{\nu_d} = R = 0.028. \tag{24}$$

Since $R \ll 1$, the density of the first excited state is essentially determined by electron collisions. The rate of excitation is

$$\nu_x = n_o n_e \int_{E_x}^{\infty} f_e(E)\, \sigma_x(E)\, v_e(E)\, dE. \tag{25}$$

It is now necessary to specify the electron energy distribution. If this distribution is Maxwellian, then the integrals in Eqs. 22 and 25 are unique functions of the electron temperature T_e. Taking the ratio of the two equations yields

$$\frac{\nu_d}{\nu_x} = \frac{n_x}{n_o} F(T_e). \tag{26}$$

The function $F(T_e)$ can be determined, since at equilibrium $\nu_x = \nu_d$ and

$$\left(\frac{n_x}{n_o}\right)_{equil} = \frac{\omega_x}{\omega_o} \exp\left(-\frac{E_x}{kT_e}\right), \tag{27}$$

so that

$$\frac{\nu_d}{\nu_x} = \frac{n_x}{n_o} \frac{\omega_o}{\omega_x} \exp\left(\frac{E_x}{kT_e}\right). \tag{28}$$

Combining Eqs. 23 and 28 yields

$$\frac{n_x}{n_o} = \frac{\omega_x}{\omega_o} \exp\left(-\frac{E_x}{kT_e}\right)\frac{1}{1+R}. \tag{29}$$

Since $R \ll 1$, the density ratio of excited to ground states is close to the equilibrium value given by Eq. 27. This ratio is plotted in Fig. XII-34.
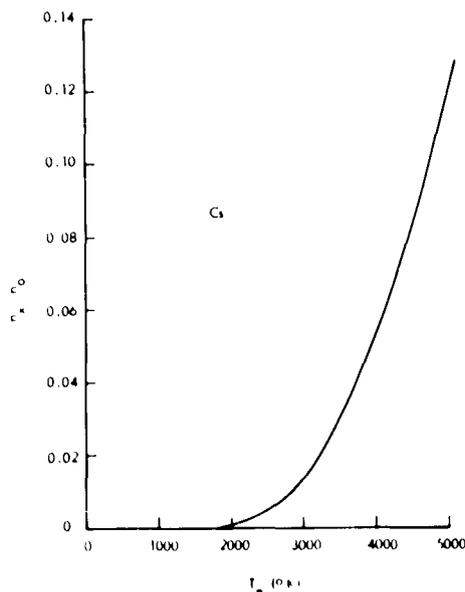
Fig. XII-34. Equilibrium density ratio of the 6p excited state and ground state of cesium plotted as a function of electron temperature.

There is a large number of other mechanisms that affect the density of excited states in a plasma, and it is virtually impossible to calculate the contribution of each of them. It has been shown in this report, however, that the inelastic electron collision cross sections are very large. Since the electron density is quite high, the inelastic electron collision rates are dominant in determining the concentration of the first excited state. This conclusion is directly applicable to the plasma in the high-current mode of a cesium thermionic-energy converter.

H. L. Witting

References

1. R. B. Brode, The absorption coefficient for slow electrons in alkali metal vapors, Phys. Rev. **34**, 673 (1929).

2. D. R. Bates, A. Fundaminsky, H. S. W. Massey, and J. W. Leech, Excitation and ionization of atoms by electron impact – the Born and Oppenheimer approximations, Phil. Trans. Roy. Soc. London **A243**, 93 (1950).

3. M. J. Seaton, The impact parameter method for electron excitation of optically allowed atomic transitions, Proc. Phys. Soc. (London) **79**, 1105 (1962).

4. W. Christoph, Ann. Phys. **23**, 51 (1935).

5. G. Haft, Z. Phys. **83**, 73 (1933).

6. R. H. Fowler, Phil. Mag. <u>47</u>, 257 (1924).

7. A. C. G. Mitchell and M. W. Zemansky, <u>Resonance</u> <u>Radiation</u> <u>and</u> <u>Excited</u> <u>Atoms</u> (Cambridge University Press, London, 1961).

8. C. Gregory, Phys. Rev. <u>61</u>, 465 (1942).

9. W. Furssow and A. Wlassow, Physik. Z. Sowjetunion <u>10</u>, 378 (1936).

10. R. W. Nichols and A. L. Stewart, Allowed transitions, <u>Atomic</u> <u>and</u> <u>Molecular</u> <u>Processes</u>, edited by D. R. Bates (Academic Press, Inc., New York, 1962), Chapter 2, p. 49.

11. A. C. G. Mitchell and M. W. Zemansky, <u>op</u>. <u>cit</u>., p. 96.

K. ELECTRON GUN FOR PRODUCTION OF A LOW-DIVERGENCE BEAM

A program has been initiated to determine experimentally the properties of the cork-screw.[1] An electron beam of low current (0.1-20 μa) and minimum divergence (less than
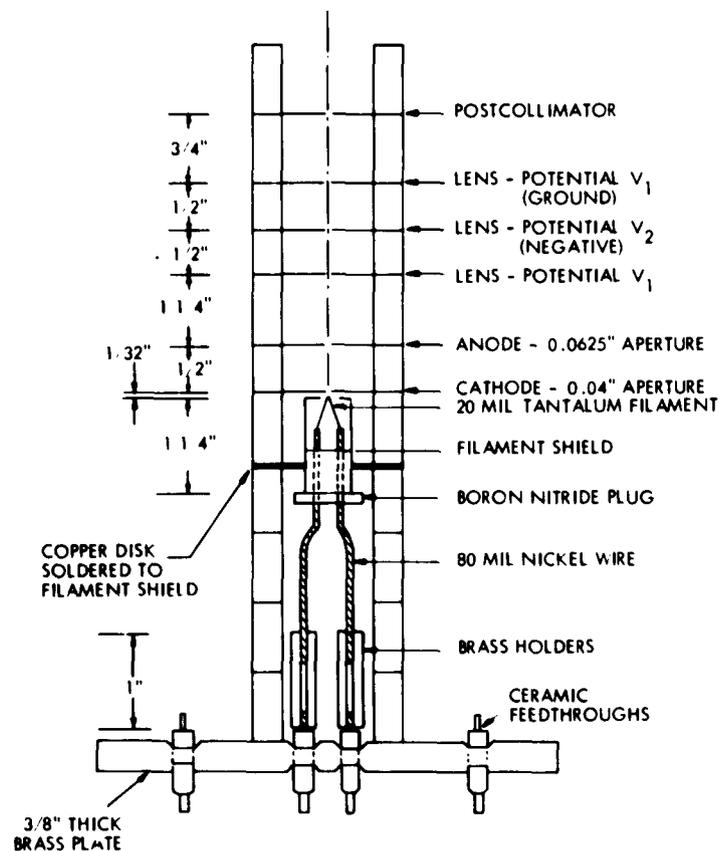
Fig. XII-35. Electron gun for the production of a low-divergence beam.

1/500 radian) will be required, and an electron gun has been constructed to produce the beam. The properties of the present model are:

Beam Current, 0.1-5 μa

Gun Length, 20 cm

Lens Aperture, 0.159-cm diameter

Cathode Aperture; 0.102-cm diameter.

The electron gun is shown in Fig. XII-35.

The beam currents investigated (0.1-3 μa) were well within the region of negligible space-charge effects, thereby giving a divergence angle of 1/1000 radian at a distance of 2000 lens aperture radii from the end of the gun for a beam current of approximately 15 μa.

In three separate trials, the average beam-divergence angles over the path length investigated (1.5 meters) were 1.8, 1.91, and 1.59×10^{-3} radian.

The observed divergences can be attributed to aberration effects in the gun; spherical aberration is the predominant cause. Alterations that will be applied to reduce the aberrations are: (a) adjustment of one relative aperture of one negative lens element, and (b) addition of postcollimators to define the beam and remove widely divergent particles.

The gun will then be scaled down by a factor of approximately four for insertion in the corkscrew experiment.

P. Karvellas

References

1. D. J. Rose, R. C. Wingerson, and S. J. Tulenko, Helical resonant beam trapping, Quarterly Progress Report No. 63, Research Laboratory of Electronics, M. I. T., October 15, 1961, pp. 41-44.

L. SPACE-CHARGE NEUTRALIZATION EXPERIMENT

The purpose of the experiment is to examine ways of injecting electrons into an ion beam in order to neutralize its space charge and avoid the consequent stalling of the beam. Practical applications of the experiment arise in spacecraft ion propulsion.

The ion beam is extracted from a cesium ion source which is shown schematically in Fig. XII-36.

The porous tungsten ionizer, a, through which the cesium atoms diffuse, is heated to approximately 1200°C by electron bombardment of its tantalum support.

The method by which electrons are brought into the region of the ion beam is by extracting them from a filament, c, through the accelerating grid, b.

The experiments were performed in a 14 in. × 15 in. × 36 in. aluminum vacuum
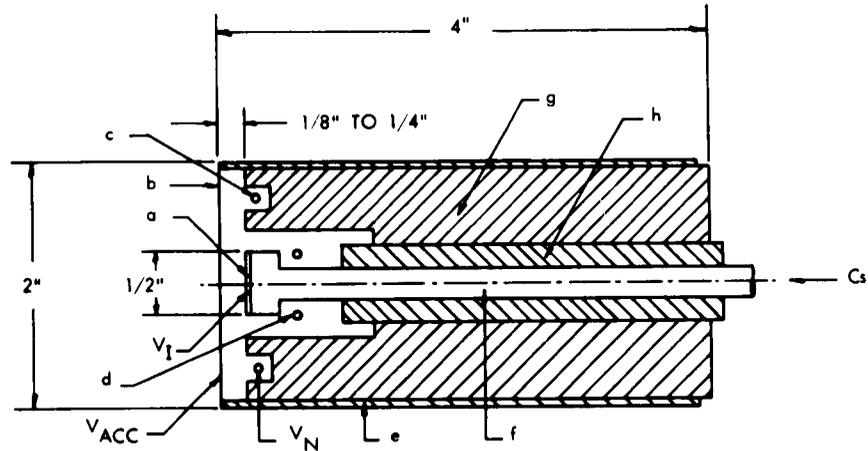
Fig. XII-36. Cesium ion source.
a. Porous tungsten ionizer
b. Accelerating grid
c. Filament producing neutralizing electrons
d. Filament for heating the ionizer
e. Accelerating electrode
f. Tantalum tube for Cs supply
g. Insulator
h. Insulator

chamber, under a pressure of approximately 2×10^{-6} mm Hg.

The current, I, was measured on a collector, 8 inches in diameter, placed 30 inches away from the source. The collector, as well as the vacuum chamber, was at ground potential. The voltage of the ionizer, the accelerating grid, and the neutralizing filament will be denoted V_I, V_{Acc}, and V_N, respectively.

Under steady-state operation of the source, currents I of from 10 ma to 20 ma at voltages, V_I, of 3000-5000 volts were obtained, without any neutralizing electrons being introduced. Space-charge considerations show that an unneutralized ion beam for such currents is impossible at these voltages. Therefore one must conclude that electrons were obtained from the environs (for example, by secondary emission from the walls). The electrons are trapped in the volume of the beam (for example, by losing energy through inelastic collisions), and in this way form a neutral path through which the ion current can pass.

This type of neutralization will not occur in space, of course. Its presence, though, in the laboratory does not allow testing of other ways of injecting electrons.

In order to overcome this difficulty, we examined the time period immediately after the ion extraction voltage is applied and before the described volume neutralization is completed.

The ionizer voltage V_I was applied as a step with a rise time of less than 0.5 μsec,
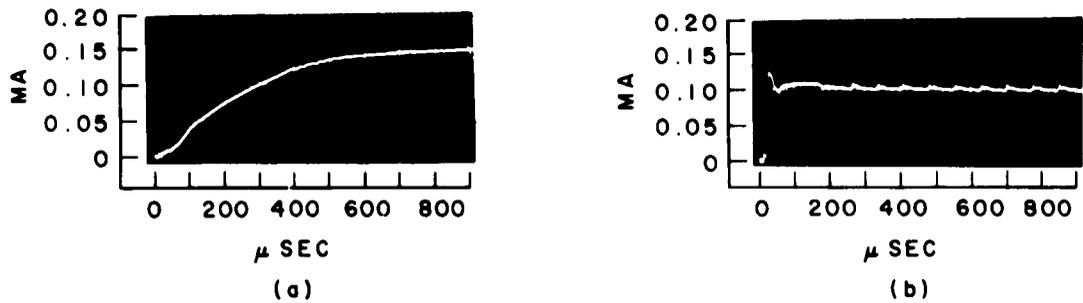
Fig. XII-37. Collected ion currents.
V_I = 600 volts, V_{Acc} = 100 volts, V_N = 0. Transit time, 22 μsec.
(a) No electrons: rise time, ~500 μsec.
(b) With electrons: rise time, ~25 μsec.
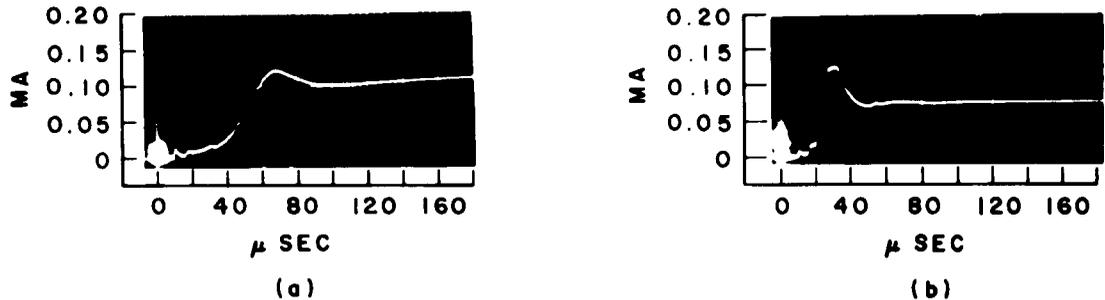
Fig. XII-38. Collected ion currents.
V_I = 1000 volts, V_{Acc} = 200 volts, V_N = 100 volts.
Transit time, 18 μsec.
(a) No electrons: rise time, ~60 μsec.
(b) With electrons: rise time, ~20 μsec.
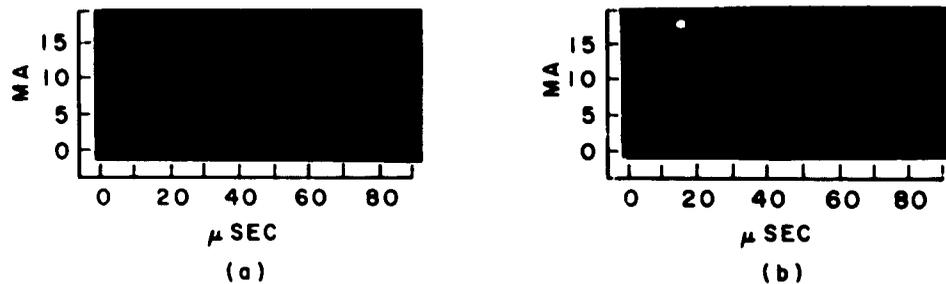
Fig. XII-39. Collected ion currents.
V_I = 4000 volts, V_{Acc} = 300 volts, V_N = 100 volts.
Transit time, 9 μsec.
(a) No electrons: rise time, ~20 μsec.
(b) With electrons: rise time, ~10 μsec.

and the collector current I was subsequently recorded on an oscilloscope.

At low voltages (V_I < 700 volts) we found that the time taken for the full ion current to reach the collector was considerably longer than the transit time for ions of this energy (Fig. XII-37a). This fact indicates that the beam is temporarily stopped by its excessive space charge until enough electrons are accumulated from the environs and trapped in the beam to achieve neutralization.

When electrons were injected from the neutralizing filament the current on the collector reached its full value in a time that was close to the ion transit time.

For example, at V_I = 600 volts and I $\approx$ 0.1 ma (Fig. XII-37), the full current reaches the collector in approximately 500 μsec without electrons, and in approximately 25 μsec with electrons. The ion transit time is 22 μsec.

Similar results have been reported by Sellen and Kemp[1] when electrons are made available to the beam from hot filaments placed in its proximity.

At higher extraction voltages and ion currents, the time taken for the beam to reach the collector when no electrons are introduced is much shorter, being only 2 or 3 times the transit time. It appears that under these conditions electrons can be obtained and trapped faster than before, probably because of the higher electric fields that are present.

Introduction of electrons reduces this time still further, bringing it to the level of the transit time. For example, at V_I = 1000 volts, I = 0.1 ma (Fig. XII-38), introduction of electrons reduces the rise time from approximately 60 μsec to approximately 20 μsec, against a transit time of 18 μsec.

At V_I = 4000 volts, I = 10 ma (Fig. XII-39) the rise time is approximately 20 μsec without electrons and approximately 10 μsec with electrons, while the transit time is 9 μsec.

These results indicate that electrons introduced in the region of the beam by an electron gun are effecti·· in neutralizing the beam space charge.

Further experi: ?nts are in progress in order to establish the exact beam pattern under various cond. ?ns. To accomplish this, the walls of the chamber have been covered with isolated metallic strips, and the collected currents individually measured.

The effect of variations in different parameters, such as the accelerating grid voltage V_{Acc}, neutralizer voltage V_N, etc., on the behavior of the beam, is also being investigated.

G. C. Theodoridis

References

1. J. M. Sellen and R. F. Kemp, Cesium Ion Beam Neutralization in Vehicular Simulation, American Rocket Society Paper 61-84-1778, presented at the IAS-ARS Joint Meeting, Los Angeles, California, June 1961.

XIII. PLASMA MAGNETOHYDRODYNAMICS AND ENERGY CONVERSION[*]

Prof. G. A. Brown
Prof. E. N. Carabateas
Prof. S. I. Freedman
Prof. W. H. Heiser
Prof. M. A. Hoffman
Prof. W. D. Jackson
Prof. J. L. Kerrebrock
Prof. H. P. Meissner
Prof. J. R. Melcher
Prof. G. C. Oates
Prof. J. P. Penhune
Prof. J. M. Reynolds III
Prof. A. H. Shapiro
Prof. J. L. Smith, Jr.
Prof. R. E. Stickney
Prof. H. H. Woodson
Dr. R. Toschi
A. A. Aponick
E. R. Babcock
M. T. Badrawi
J. F. Carson
A. N. Chandra

R. S. Cooper
J. M. Crowley
R. Dethlefsen
M. G. A. Drouet
D. A. East
J. R. Ellis, Jr.
M. R. Epstein
F. W. Fraim IV
J. W. Gadzuk
J. Gerstmann
N. Gothard
J. B. Heywood
R. M. Jansen
H. D. Jordan
P. G. Katona
F. D. Ketterer
G. B. Kliman
A. G. F. Kniazzeh
H. C. Koons
M. F. Koskinen
K. S. Lee
R. F. Lercari
W. H. Levison

A. T. Lewis
H. C. McClees, Jr.
C. W. Marble
T. D. Masek
J. T. Musselwhite
S. A. Okercke
J. H. Olsen
C. R. Phipps, Jr.
E. S. Pierson
D. H. Pruslin
M. H. Reid
C. W. Rook, Jr.
A. W. Rowe
M. R. Sarraquigne
A. Shavit
M. N. Shroff
A. Solbes
P. M. Spira
J. S. Weingrad
G. L. Wilson
J. C. Wissmiller
B. M. Zuckerman

A. WORK COMPLETED

1. A-C PROPERTIES OF SUPERCONDUCTORS

The present phase of this work has been completed by C. R. Phipps and the results have been presented in a thesis entitled "Alternating-Current Properties of Superconducting Wires" to the Department of Electrical Engineering, M.I.T., in partial fulfillment of the requirements for the degree of Master of Science, June 1963.

W. D. Jackson

2. HYDROMAGNETIC WAVEGUIDES

The present phase of this work has been completed by M. R. Epstein and the results have been presented in a thesis entitled "Alfvén Wave Propagation in Liquid NaK" to the Department of Electrical Engineering, M.I.T., in partial fulfillment of the requirements for the degree of Bachelor of Science, June 1963.

W. D. Jackson

3. MAGNETOHYDRODYNAMIC VORTEX GENERATOR

This research has been completed by H. D. Jordan and the results have been submitted as an S. B. thesis entitled "An Evaluation of the Vortex-Type Magnetohydrodynamic Generator" to the Department of Electrical Engineering, M. I. T., June 1963.

W. D. Jackson

4. IONIC PLASMA OSCILLATIONS

The present phase of this work has been completed by H. C. Koons and the results have been submitted as an S. B. thesis entitled "Ultrasonic Excitation of Ionic Plasma Oscillations" to the Department of Physics, M. I. T., June 1963.

R. S. Cooper

5. FLOWMETERS FOR BLOOD-FLOW MEASUREMENT[†]

The present phase of this work has been completed and the results have been presented as theses to the Department of Electrical Engineering, M. I. T.

C. W. Marble, "A Thermistor Equipped Thermal Conduction Flowmeter," S. B. Thesis. June 1963.

J. T. Musslewhite, "Design of an Electromagnetic Flowmeter," S. B. Thesis, June 1963.

M. N. Shroff, "Calibration of Electromagnetic Flowmeter for Blood Flow Measurement," S. B. Thesis, June 1963.

W. D. Jackson

6. BLOOD-FLOW STUDIES[†]

The present phase of this work has been completed by M. R. Sarraquigne and the results have been presented in a thesis entitled "A Pressure Function Generator for Blood Flow Studies" to the Department of Electrical Engineering. M. I. T., in partial fulfillment of the requirements for the degree of Bachelor of Science, June 1963.

W. D. Jackson

B. A-C PROPERTIES OF SUPERCONDUCTORS

Investigation of the properties of superconducting coils under alternating current excitation is being pursued to determine the feasibility of using available superconductors

---

to fabricate very low loss ac magnetic fields. Both hard and soft superconductors are being tested from dc up to a frequency of 10 kc.

Two quantities characterizing the ac properties of these materials have been measured.

(a) $I_q(\omega)$ the maximum instantaneous supercurrent that can be applied instantaneously.

(b) $I_c(\omega)$ the maximum supercurrent that can be maintained indefinitely without the occurrence of a quench.

The frequency dependence of these two quantities has been found to differ in hard and soft superconductors. Results of tests on solenoids wound with Nb-Zr and lead wires are given in Fig. XIII-1. The procedure for establishing values of I_q was to
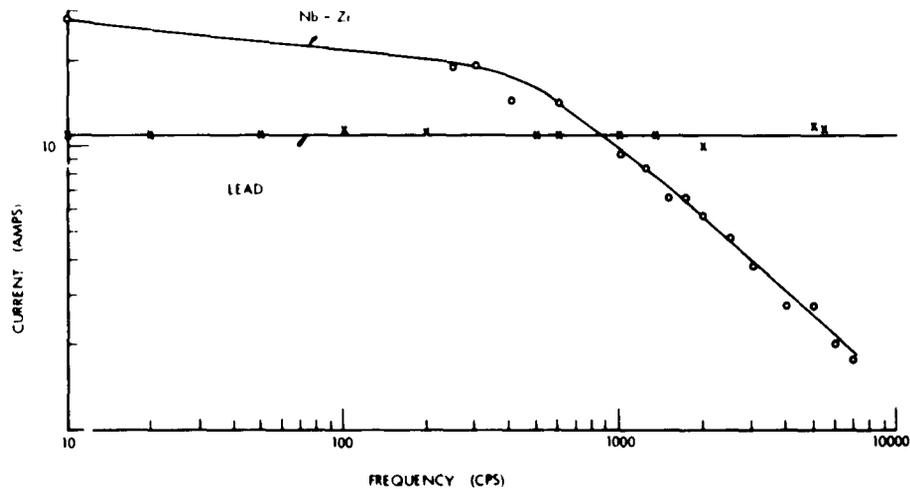
Fig. XIII-1. $I_q(\omega)$ for Nb-Zr and for lead.

increase the solenoid current in approximately 1 second to a value that would cause a rapid quench. By repeated trials, the I_q to obtain a quench in less than 0.5 sec was determined.

Figure XIII-1 shows $I_q(\omega)$ to be independent of frequency in lead, the soft superconductor tested, up to 10 kc, while for Nb-Zr, a hard superconductor, the drop-off in $I_q(\omega)$ with increasing frequency is such that at 7 kc, $I_q(\omega)$ is approximately 7 per cent of the 10-cps value. Hard superconductors are assumed to carry current in internal filaments, whereas soft superconductors carry it in a thin skin at the surface. Changes in $I_q(\omega)$ appear to be controlled by frequency dependence of the filamentary structure, the detailed nature of which is still inknown.

In both hard and soft superconductors, time-delayed quenches could be attained; thus a difference between I_c and I_q is indicated. As illustrated in Figs. XIII-2 and XIII-3,
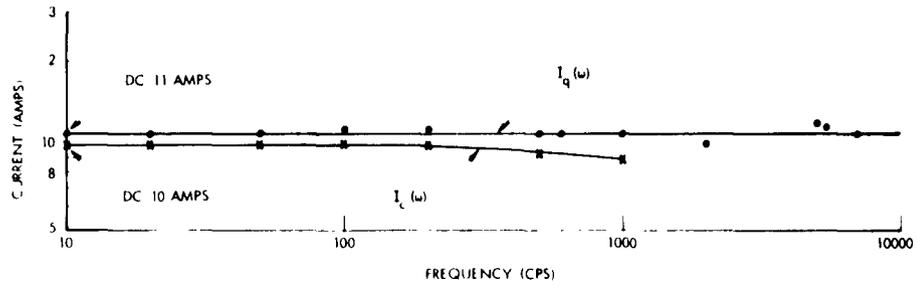
Fig. XIII-2. $I_c(\omega)$ and $I_q(\omega)$ for lead.
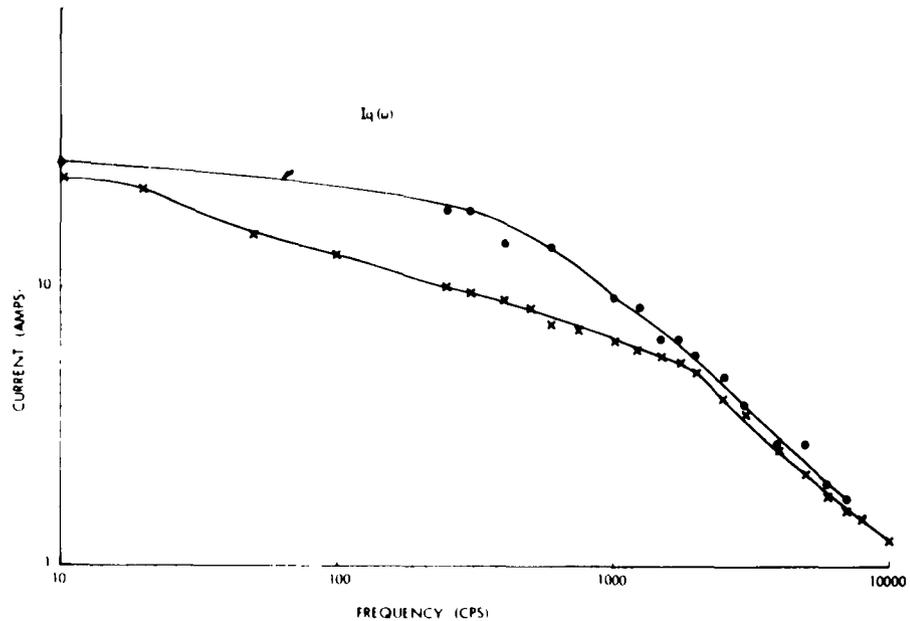
Fig. XIII-3. $I_q(\omega)$ and $I_c(\omega)$ for Nb-Zr.

these occurred in Nb-Zr at much lower frequencies than in lead. These quenches are apparently due to ac loss which heats the specimen to a temperature higher than its critical temperature. They should be more pronounced in hard superconductors, which have an internal field in the ohmic region, and this has been confirmed by experimental observation. In soft superconductors such losses occur in the surface region only.

This note is based on two recent theses[1,2] presented to the Department of Electrical Engineering, M. I. T. Further work is contemplated to confirm and extend the results thus far obtained.

W. D. Jackson, C. R. Phipps, Jr., P. M. Spira

References

1. C. R. Phipps, Alternating-Current Properties of Superconducting Wires, S. M. Thesis, Department of Electrical Engineering, M. I. T., June 1963.

2. P. M. Spira, Superconducting Coil Properties with Alternating Current Excitations, S. B. Thesis, Department of Electrical Engineering, M. I. T., June 1963.

C. LANGMUIR-MODE ANALYSIS OF THE PLASMA THERMIONIC-ENERGY CONVERTER

1. Introduction

Under certain conditions of emitter temperature and cesium pressure the voltage-current characteristic of a cesium thermionic converter exhibits two branches. The upper branch, at higher output power and efficiency, is considered as a Langmuir-mode gaseous electric discharge. The Langmuir-mode discharge[1] is characterized by a neutral plasma joined to the emitter by a double sheath and to the collector with a unipolar sheath.

Ions are created in the interelectrode spacing by inelastic collisions of the electrons. These ions diffuse to the emitter and collector, and thus decrease the potential barrier in front of the emitter and permit a larger electron current to flow.

In the present report an analysis of the Langmuir-mode discharge based on the following model is carried out.

Electrons are emitted from the cathode at A on the motive diagram of Fig. XIII-4, pass over C (the top of the potential barrier), and fall through the thin double sheath CD and the presheath DE, and finally enter the cesium plasma at E. On entering the plasma the electron velocities are immediately randomized by collisions with Cs atoms and Maxwellianized by collisions with each other. The electron current spilling over the barrier is limited by the ion current falling from the plasma back to the cathode according to the Langmuir double-sheath condition.[2] Ions are generated in the plasma by a cumulative process, and diffuse to the electrodes by ambipolar diffusion. The total ionization probability is determined to be an eigenvalue to the ion-diffusion equation and the electron temperature is found by satisfying the electron-energy equation. These equations are sufficient to permit calculation of I-V curves (as shown in Fig. XIII-9) for different values of p_o the reduced cesium pressure, d the diode spacing, T_E the emitter temperature, and T_C the collector temperature. These calculations were performed by choosing two constants pertaining to the excitation and ionization rates[3] so that the theoretical and experimental curves would coincide at one data point. The derivation is presented below.

2. Diffusion Equation

Transport in the plasma is by ambipolar diffusion with a charge-transfer cross section.[4]

$$j_p = -D_a \nabla n_e - \frac{\mu_p}{\mu_e} j_e$$

$$\nabla \eta = -\frac{\nabla n_e}{n_e} - \frac{j_e}{n_e D_e}$$

$$\eta \equiv \frac{eV}{kTe} \qquad D_a = \frac{3\pi\sqrt{2}}{8} \lambda_p \frac{\bar{c}_p}{4} \left(\frac{T_e}{T_p} + 1 \right)$$

(1)

$$\lambda_p = \frac{1}{9300 p_o} \qquad \bar{c}_p \equiv \sqrt{\frac{8kT_p}{\pi M}}$$

Excited 6p atoms are generated by electron-atom collisions at a rate

$$\frac{dn^*}{dt} = a_1 n_e$$

where

$$a_1 = \alpha_1 p_o F_1$$

$$F_1 = 1.97 \times 10^7 \left(\frac{T_e}{10^3} \right)^{1/2} \left(1.39 + \frac{T_e}{5800} \right) \exp\left(-\frac{16.15 \times 10^3}{T_e} \right).$$

Ions are generated by electron-atom collisions at a rate

$$\frac{dn_e}{dt} = a_2 \frac{n^*}{n_a} n_e$$

where

$$a_2 = \alpha_2 p_o F_2$$

$$F_2 = 1.97 \times 10^7 \left(\frac{T_e}{10^3} \right)^{1/2} \left(2.5 + \frac{T_e}{5800} \right) \exp\left(-\frac{29 \times 10^3}{T_e} \right).$$

Excited atoms are lost by imprisoned resonance radiation having an effective lifetime $\tau = 2.55 \times 10^6$ sec. In a steady state

$$a_1 n_e = \frac{n^*}{\tau}.$$

Ions are lost by diffusion

$$-D_a \nabla^2 n_e = \nabla \cdot j_p = a_2 \frac{n^*}{n_a} n_e = a_1 a_2 \tau \frac{n_e^2}{n_a}.$$

In one dimension,

$$\frac{d^2 z}{d\xi^2} = -\frac{3}{2} z^2$$

$$z \equiv \frac{n_e}{n_{eo}} \qquad \xi \equiv \frac{x}{\ell} \tag{2}$$

$$\ell^2 \equiv \frac{3}{2} \frac{D_a}{a_1 a_2 \tau \frac{n_{eo}}{n_a}}.$$

The diffusion equation is solved by using the transformation

$$z = \cos^{2/3} \theta \qquad \frac{dz}{d\xi} = -\sin \theta$$

$$\frac{d\theta}{d\xi} = \frac{3}{2} \cos^{1/3} \theta \qquad \xi \approx 1.403 - \cos^{2/3} \theta.$$

At the plasma boundary the ion current equals the random-ion current.

$$j_p = n_e \frac{\bar{c}_p}{4}.$$

At the emitter edge,

$$\cos^{2/3} \theta_E = \frac{4.66 \frac{\lambda_p}{d} \left(\frac{T_e}{T_p} + 1 \right)}{\left[1 + 3.32 \frac{\lambda_p}{d} \left(\frac{T_e}{T_p} + 1 \right) \right] (1-R)}.$$

At the collector edge,

$$\cos^{2/3} \theta_C = (1-2R) \cos^{2/3} \theta_E$$

where

$$R \equiv \frac{\mu_p}{\mu_e} \frac{j_e}{j_{pE}}.$$

The mobilities for ions and electrons are evaluated[5,6] to be

$$\mu_p = \frac{e}{kT_p}\, \frac{3\pi\sqrt{2}}{8}\, \lambda_p\, \frac{\bar{c}_p}{4} \qquad \mu_e = \frac{e}{kT_p}\, \frac{1}{3}\, \lambda_e\, \bar{c}_e.$$

$$\frac{\lambda_p}{\lambda_e} = \frac{P_C(e)}{P_C(p)} = \frac{1200}{9300}.$$

Therefore

$$R = 0.161 \left(\frac{T_e}{T_p}\right)^{1/2} G$$

where

$$G \equiv \left(\frac{m}{M}\right)^{1/2} \frac{j_e}{j_{pE}}.$$

It will be shown that $G \leq 1$, so that R is always small.

The ion currents to the emitter and collector are

$$j_{pE} = \frac{n_{eo}}{\ell}\, D_a (1-R)^{-1}$$

$$j_{pC} = (1-2R)\, j_{pE}.$$

The potential drop in the plasma (EF in Fig. XIII-4) is determined by integrating Eq. 1 for the reduced field. For small R the result is

$$\eta_e - \eta_a \approx \frac{G}{1-R}\left[1.54 - 0.14\,\frac{T_e}{10^3} + \left(0.5\,4 + 0.050\,\frac{T_e}{10^3}\right)\ln\left(\frac{P_C P_o d}{100}\right)\right]. \tag{3}$$

3. Electron-Energy Equation

Electrons enter the plasma with an average kinetic energy corresponding to their original heat of transport plus the energy gained in dropping down the cathode fall (CE in Fig. XIII-4). They leave the plasma with a different energy, corresponding to the energy necessary to traverse the barrier GH plus the heat of transport. While in the plasma, the electrons lose energy by exciting and ionizing atoms. In the steady state the net energy carried into the plasma by electrons must equal the energy carried out by radiation and ion transport. These various terms are

Radiation: $\int e(1.39)\frac{n^*}{\tau}\,dx = \int e(1.39)\,a_1 n_e\,dx = 0.305\,\frac{a_1 d^2}{D_a}\,ej_{pE}(1-R)$.

Ion transport: $e(3.89)(j_{pE}+j_{pC}) = 7.78ej_{pE}(1-R)$.

Net electron transport in: $ej_e\left[\Delta V - \frac{2k}{e}(T_e-T_E)\right] - ej_b\,\frac{2k}{e}(T_e-T_E)$

$$- ej_{pE}(1-R)\,2\left(\frac{2kT_e}{e} + V_a\right).$$

The first term in the net electron transport is due to the net electron current j_e; ΔV is the effective plasma drop CH. The second term is due to transport of energy to the cathode which, in turn, is due to the current emitted back from the plasma j_b; the third

A	CATHODE FERMI LEVEL
I	ANODE FERMI LEVEL
AI	OUTPUT VOLTAGE, $\frac{V}{.T_e}$
CH	EFFECTIVE PLASMA DROP
AB	CATHODE WORK FUNCTION
AC	EFFECTIVE CATHODE WORK FUNCTION
CD	LANGMUIR DOUBLE SHEATH POTENTIAL
DE	CATHODE PRESHEATH
EF	PLASMA DROP
FG	ANODE PRESHEATH
GH	ANODE SHEATH
FH	EFFECTIVE ANODE SHEATH
HI	ANODE WORK FUNCTION
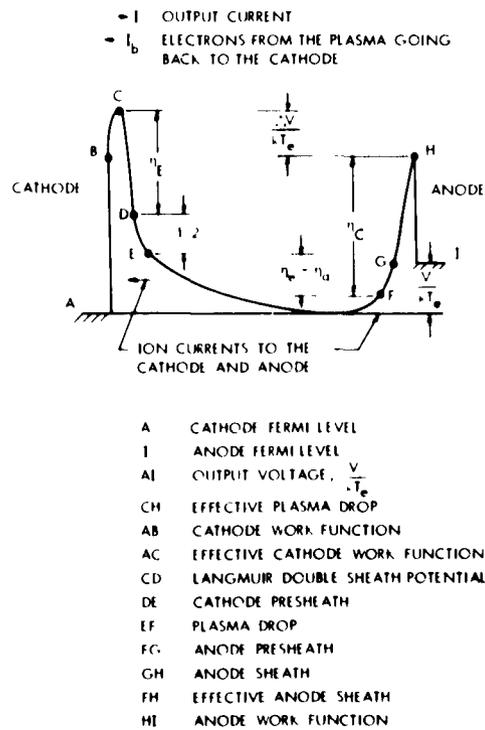
Fig. XIII-4. Motive diagram for the Langmuir mode.

term accounts for the fact that the electron current at the anode differs from that at the cathode by the net ion production in the plasma. It is assumed that an electron makes its last collision at the edge of the plasma (F in Fig. XIII-4) on its way to the anode; V_a is the effective anode sheath height (HF in Fig. XIII-4).

Equating the energy sources and losses for the plasma electrons gives

$$\eta_1 \frac{G}{1-R} = 3.98 \times 10^{-7} \, \alpha_1 P_C (P_0 d)^2 \frac{F_1}{\frac{T_e}{10^3}\left(\frac{T_e}{T_p}+1\right)}, \tag{4}$$

where

$$\eta_1 \equiv \frac{\Delta V}{kT_e} - 2\left(1 - \frac{T_E}{T_e}\right) - 2\left(\frac{m}{M}\right)^{1/2}\left(\frac{3.89 \times 11,605}{T_e} + \frac{V_a}{kT_e} + 2\right)\frac{(1-R)}{G} - 2\left(1 - \frac{T_E}{T_e}\right)\frac{j_b}{j_e}. \tag{5}$$

The relation between G and η_1 (defined below) provides the last equation required for the calculation of the voltage-current curves for the diode.

4. Langmuir Double-Sheath Condition

The last important relation is derived by integrating Poisson's equation across the cathode double sheath. With the conditions that the electric field is zero at the potential maximum and at the plasma edge of the sheath (C and D of Fig. XIII-4), and that the charge density is zero at D, one finds a relation[2] between the sheath voltage and the ratio of the electron and ion currents passing through the sheath.

$$G = G\left(\eta_E, \frac{T_E}{T_e}, \frac{T_{p\,eff}}{T_e}\right), \tag{6}$$

where η_E is the reduced voltage corresponding to CD in Fig. XIII-4.

The Bohm criterion[7] for the formation of a nonoscillatory sheath fixes a limit on $T_{p\,eff}$, the effective temperature of the ions leaving the presheath: $T_{p\,eff}/T_e \geq 0.5$. G is plotted as a function of η_E for various T_E/T_e and for $T_{p\,eff}/T_e = 0.5$ in Fig. XIII-5. G approaches 1 as η_E becomes very large but is usually approximately 0.5.

The Bohm criterion also predicts the presheath voltage (DE in Fig. XIII-4) to be $0.5kT_e$.

The Langmuir condition also relates the back current and the forward current across the sheath:

$$\frac{j_b}{j_e} = \left[\frac{\sqrt{2}}{G} - \frac{1}{\sqrt{\eta_E + \frac{T_E}{T_e}}}\right]\left[2\sqrt{\pi}\,\exp\eta_E + \frac{1}{\sqrt{\eta_E + \frac{T_E}{T_e}}} - \frac{1}{\sqrt{\eta_E + 1}}\right]^{-1}. \tag{7}$$

The height of the unipolar anode sheath is given by the Boltzmann factor

$$\exp\eta_C(1+\text{erf}\sqrt{C}) = 2\left(\frac{T_e}{T_p}\right)^{1/2}\frac{(1-R)}{G}. \tag{8}$$
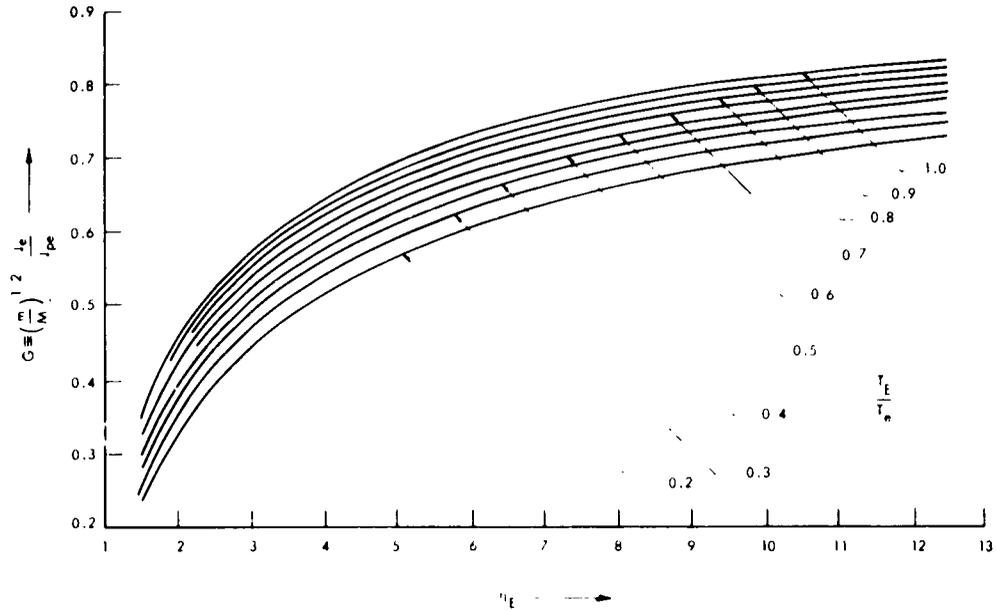
Fig. XIII-5. Langmuir double-sheath condition $T_{p\ eff} = 0.5T_e$.

Equations 3 and 5-8 are combined to give $\dfrac{G}{1-R}$ as a function of

$$\eta_2 \equiv \eta_1 - \frac{G}{(1-R)} \left(0.564 + 0.050 \frac{T_e}{10^3}\right) \ln\left(\frac{P_C P_o d}{100}\right),$$

as shown in Fig. XIII-6. These curves are approximated by straight lines to give

$$\frac{G}{1-R} = \frac{0.255 + 0.041 \dfrac{T_e}{10^3} + 0.103\eta_1}{1 + \left(0.057 + 0.005 \dfrac{T_e}{10^3}\right) \ln\left(\dfrac{P_C P_o d}{100}\right)} \qquad \text{if } \frac{G}{1-R} < 0.57 \qquad (9)$$

$$\frac{G}{1-R} = \frac{0.387 + 0.02 \dfrac{T_e}{10^3} + 0.065\eta_1}{1 + \left(0.037 + 0.0032 \dfrac{T_e}{10^3}\right) \ln\left(\dfrac{P_C P_o d}{100}\right)} \qquad \text{if } \frac{G}{1-R} > 0.57$$
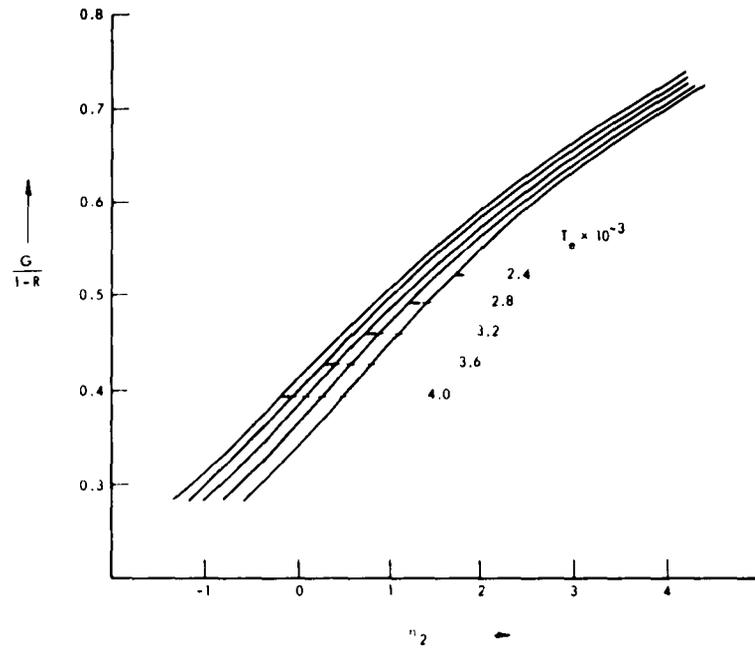
Fig. XIII-6. $\dfrac{G}{1-R}(\eta_2)$ for $T_E = 1600°K$, $T_e = 1200°K$.

5. Voltage-Current Relation

Equations 4 and 9 are combined to give a noniterative procedure for generating the voltage-current curves. The energy equation is

$$\eta_1 g = 0.0397 a_1 P_C (P_o d)^2 \ f \ \frac{F_1}{10^2 T_e \left(\dfrac{T_e}{T_p} + 1\right)}, \tag{10}$$

where

$$g = 0.255 + 0.041 T_e/10^3 + 0.103 \eta_1$$

$$f = 1 + 0.072 \ln\left(\frac{P_C P_o d}{100}\right).$$

Evaluating l from the boundary conditions and using Eq. 2 gives

$$\eta_1 I = \frac{0.122}{a_2 \tau P_C P_o d} \frac{10^8 \left(\dfrac{T_e}{T_p} + 1\right)}{F_2 T_e}. \tag{11}$$

Here, the temperature function is approximated by

$$\frac{F_2 T_e}{10^8 \left(\dfrac{T_e}{T_p} + 1\right)} = 1.3 \left[\frac{F_1}{10^2 T_e \left(\dfrac{T_e}{T_p} + 1\right)}\right]^{2.32} \cdot \tag{12}$$
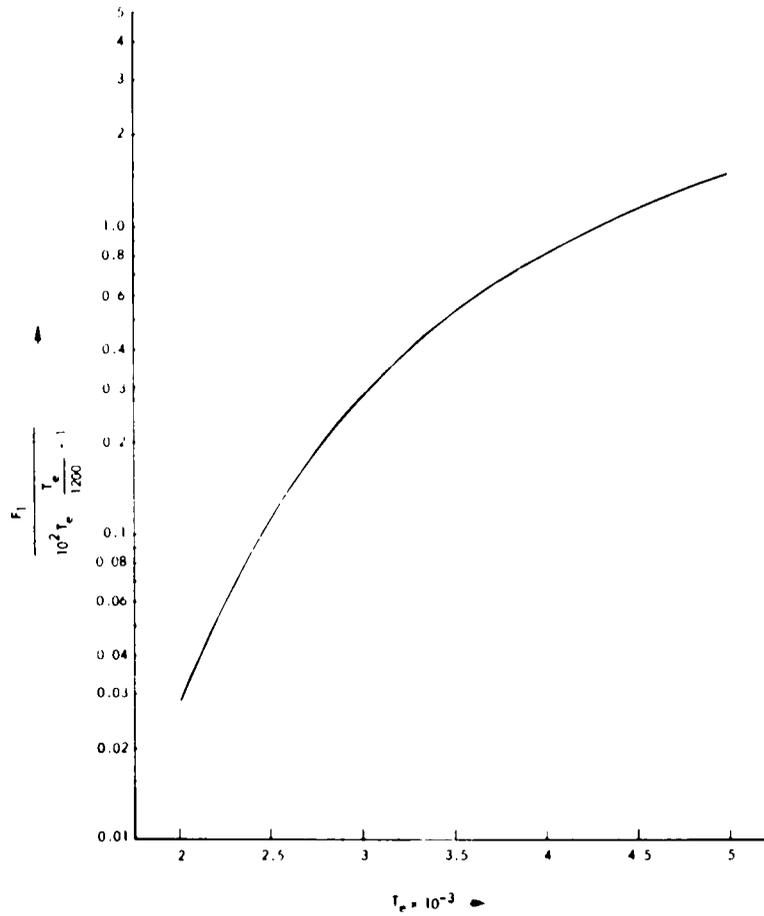
Fig. XIII-7. $\dfrac{F_1}{10^2 T_e \left(\dfrac{T_e}{1200} + 1\right)}$ versus T_e.

Equations 10-12 give

$$\eta_1 = AC^{-.7} \tag{13}$$

$$\frac{F_1}{10^2 T_e \left(\dfrac{T_e}{T_p} + 1\right)} = BC^{+.3} \tag{14}$$

where

$$A \equiv 0.0516 \frac{a_1^{.7}}{(a_2\tau)^{.3}} (p_o d)^{1.1} P_C^{.4} \left[1 + 0.072 \ln\left(\frac{P_C p_o d}{100}\right) \right]^{.3} I^{-.3}$$

$$B \equiv 1.3 \left[I a_1 a_2 \tau \left\{ 1 + 0.072 \ln\left(\frac{P_C p_o d}{100}\right) \right\} \right]^{-.3} (p_o d)^{-.9} P_C^{-.6}$$

$$C \equiv 0.353 + 0.069B + 0.169A.$$

In Fig. XIII-7, $F_1/10^2 T_e \left(\frac{T_e}{T_p} + 1\right)$ is plotted for $T_p = 1200°K$. The back current and ion loss term is plotted in Fig. XIII-8. For a given value of I, the output current in amp/cm^2, η_1, and T_e are found from Eqs. 13 and 14 and Fig. XIII-7. G/(1-R) is found

$$2\left(1 - \frac{1600}{T_e}\right)\frac{J_b}{J_o} + 2\left(\frac{m}{M}\right)^{1/2}\left(\frac{3.89}{kT_e} + 2 + \eta_c\right)\left(\frac{1-R}{G}\right)$$
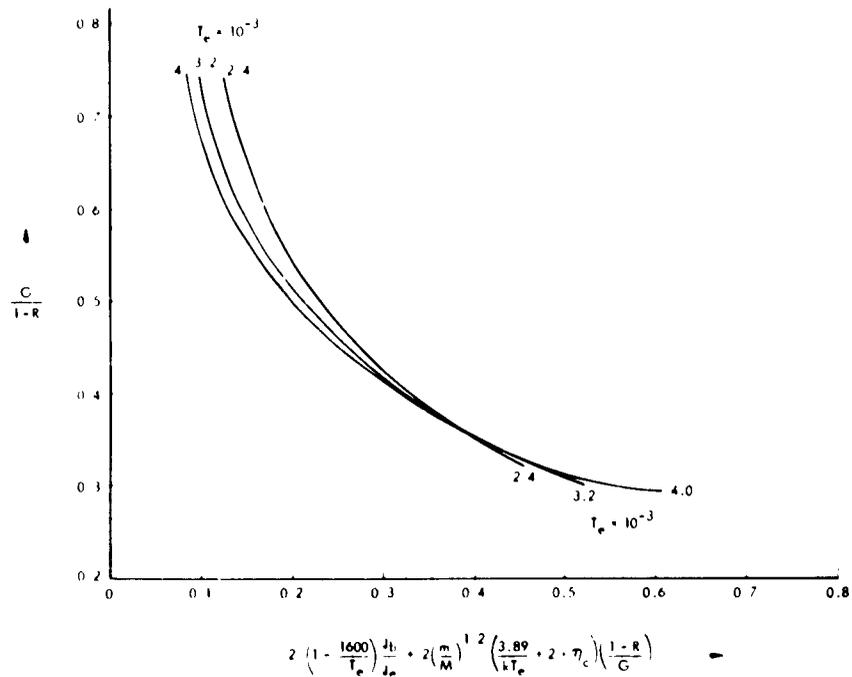
Fig. XIII-8. Back emission and ionization losses.
$T_E = 1600°K$, $T_p = 1200°K$.

from Fig. XIII-8. This is added to η_1, along with $2\left(1 - \frac{T_E}{T_e}\right)$, to give $\frac{\Delta V}{kT_e}$ and then ΔV.
Now that I and ΔV are known, the output voltage V can be found if ϕ_C, the collector

work function, is known:

$$V = \phi_m - \phi_C - V,$$

where ϕ_m is found from I and T_E by Richardson's equation.[8] Thus for given values of the diode parameters d, p_o, T_E, and ϕ_C the voltage-current curve can be found.

Theoretical I-V curves were calculated in this way and are plotted in Fig. XIII-9, along with experimental curves for the same conditions. Since accurate values of the excitation and ionization coefficients a_1 and a_2 are not available, the calculation was made by using the a_1 and a_2 that fit experimental data at one data point (shown in Fig. XIII-9) at an assumed electron temperature of 3000°K and a collector work function of 1.6 volts. This procedure results in $a_1 = 4.77$ and $a_2 = 691$.
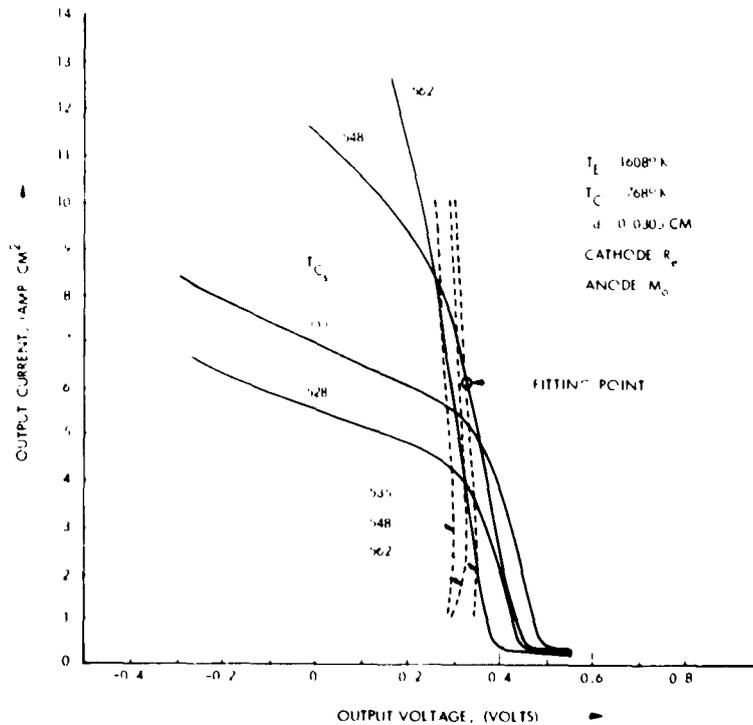
Fig. XIII-9. I-V curves from experiment (solid curves) and from theory (dashed curves, for the Langmuir-mode model).

Although these a's disagree by a factor of 7 with other estimates, the voltage-current curves derived from the model bear close resemblance to the experimental curves. The experimental curves bend over at high currents when the diode draws saturation current

from part of the emitter surface; this effect was not included in the model. The pressure dependence of the output voltage at constant current is close to that expected, although the reversal should occur at a lower cesium pressure approximately 535°K instead of 548°K. This suggests that the assumed plasma electron temperature at the fitting point is low. The slopes of the theoretical curves are not as great as the experimental ones, although some improvement would be evident if the ion current and the back current were not neglected in calculating ϕ_m from Richardson's equation.

In conclusion, it is felt that a model of the Langmuir-mode discharge can predict the essential features of experimental voltage-current curves of a thermionic diode. Several improvements of the model are suggested by this analysis: consideration of de-excitation by electron-excited atom collisions, recombination, alternative ionization mechanisms, and energy exchange by elastic recoil.

<div align="right">A. G. F. Kniazzeh, E. N. Carabateas</div>

References

1. E. O. Johnson, RCA Review 16, 498 (1955).

2. I. Langmuir, Phys. Rev. 33, 954 (1929).

3. Thermo-Electron Engineering Corporation unpublished plasma thermionic diode data, Rhenium tube #2, Run 14 (1963).

4. J. W. Sheldon, J. Appl. Phys. 34, 444 (1963).

5. S. Chapman and T. G. Cowling, Mathematical Theory of Non-Uniform Gases (Cambridge University Press, London, 1960).

6. S. C. Brown, Basic Data of Plasma Physics (The Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1959).

7. D. Bohm, Characteristics of Electrical Discharges in Magnetic Fields, edited by A. Guthrie and R. K. Wakerling (McGraw-Hill Book Company, Inc., New York, 1949), Chapter 3.

8. W. T. Norris, Thermionic Emission From a Single Crystal of Tantalum Exposed to Cesium Vapor, Sc. D. Thesis, Department of Mechanical Engineering, M. I. T. August 1962.

D. CALCULATION OF MONOPOLE WORK

It has been shown previously[1] that the work function ϕ of a conductor can be written

$$\phi = M + D - \zeta, \tag{1}$$

where D (called dipole work) is the electronic charge times the difference in electrostatic potential between the inside and the outside of the conductor, ζ is the electron degeneracy energy, and M is defined by Eq. 1. It was suggested that M could be identified with a sort of "image-force work" called monopole work. This report shows a

simple method of estimating the monopole work.

When the distance x of an electron from the plane of effective charge is great, the monopole force is given by the ideal image force $F(x)$. Therefore, the monopole work will be written

$$M = - \int_0^\infty g(x) \, F(x) \, dx,$$

where $g(x)$ is determined by the nature of the material in question. When x is larger than a few atomic radii, $g(x)$ equals unity. When x approaches zero, $g(x)$ approaches zero.

The difference in monopole work ΔM_{12} between two geometrically different conductors is

$$\Delta M_{12} = \int_0^\infty g(x) \, F_1(x) \, dx - \int_0^\infty g(x) \, F_2(x) \, dx,$$

where $F_1(x)$ and $F_2(x)$ are the associated image forces computed on the basis of ideal conductors. Note that for small x, $F_1(x)$ approximately equals $F_2(x)$, that is, $F(x)$ for a semi-infinite solid. ΔM_{12} can also be written

$$\Delta M_{12} = \int_0^\delta g(x) \, [F_1(x) - F_2(x)] \, dx + \int_\delta^\infty g(x) \, [F_1(x) - F_2(x)] \, dx,$$

where δ is a constant. If δ is selected to be small enough, the first integral is very nearly zero and can be replaced by

$$\int_0^\delta [F_1(x) - F_2(x)] \, dx,$$

which is also very nearly zero. If, at the same time, δ can be taken to be large enough so that $g(\delta)$ is very nearly unity, then $g(x)$ can be eliminated from the second integral, and ΔM_{12} is written

$$\Delta M_{12} = \int_0^\infty [F_1(x) - F_2(x)] \, dx \tag{2}$$

subject to the two conditions; there exists a distance δ such that

$$g(\delta) \approx 1 \tag{3}$$

and

$$F_1(\delta) \approx F_2(\delta). \tag{4}$$

To meet the first condition, δ should be greater than a few atomic diameters. The

second condition is met when δ is small in comparison with the curvature of the surface.

Consider the difference in monopole work between a semi-infinite solid and an ungrounded sphere of radius a. By the method of images,

$$F_1(x) = -\frac{e^2}{4\pi\epsilon_0}\frac{1}{4x}$$

$$F_2(x) = -\frac{e^2}{4\pi\epsilon_0}\left[\frac{a/r}{(r-a^2/r)^2} + \frac{1-a/r}{r^2}\right],$$

where e is the electronic charge and $r = a + x$. To evaluate the integral of Eq. 2, integrate $F_1(x)$ and $F_2(x)$ separately from $x = \eta$ to $x = \infty$, take the difference, and then let η go to zero, being careful not to drop the $(-1/8a)$ term from the first part of $F_2(x)$. The result is

$$\Delta M_{12} = \frac{e^2}{4\pi\epsilon_0}\frac{3}{8a} = 5.4/a, \tag{5}$$

where ΔM_{12} is in electron volts and a is in angstroms. That is, the monopole work of a small ungrounded sphere is slightly greater than that of a semi-infinite solid (0.01 ev for a 1000 Å sphere).

By Eq. 1, if one assumes that D and ζ are not functions of crystal size, the work function (more precisely, the first ionization potential) V_a of a small sphere of radius a is given by

$$V_a = \phi_0 + 3e^2/16\pi\epsilon_0 a, \tag{6}$$

where ϕ_0 is the work function of a semi-infinite solid.

Rowe and Kerrebrock (Section XIII-F) have applied this formula to the ionization potential of alkali-metal droplets in order to compute the electrical conductivity of the vapor. In Rowe's case, for at least 100 atoms per droplet, conditions (3) and (4) are reasonably well met. Quantum-mechanical calculations by Bardeen[2] based on a model which takes the structure of the metal into account indicate that the force on an electron outside the surface of a metal is just the classical image force. By applying Eq. 6 to a single atom, Table XIII-1 shows the computed work function ϕ_0 of a semi-infinite solid, obtained from the first ionization potential[3] V_a of the free atom and the atomic radius a.[3] The measured work function[4] ϕ is given, as is the discrepancy $\phi - \phi_0$. The explanation of this discrepancy is still not complete, but a partial explanation can be given.

For the alkali metals (I): Since the kinetic energy of the valence electron is approximately equal to the electron-degeneracy of the solid, the assumption that $\Delta\zeta = 0$ seems reasonable. Having only one valence electron, the free atom has zero "surface double layer." Therefore, $\phi - \phi_0$ should equal the dipole work D of the solid metal, which is

Table XIII-1. Work function ϕ_O of the elements, computed from Eq. 6, compared with the reported values of the work function ϕ.

	Va^a (ev)	a^a (Å)	ΔM_{12}^b (ev)	ϕ_O^c (ev)	ϕ^d (ev)	$\phi - \phi_O$ (ev)
I						
Lithium	5.40	1.55	3.48	1.92	2.49	0.57
Sodium	5.14	1.90	2.84	2.30	2.28	-0.02
Potassium	4.34	2.35	2.30	2.04	2.24	0.20
Rubidium	4.17	2.48	2.18	1.99	2.09	0.10
Cesium	3.89	2.67	2.02	1.87	1.81	-0.06
II						
Beryllium	9.32	1.12	4.82	4.50	3.92	-0.58
Magnesium	7.64	1.60	3.37	4.27	3.68	-0.59
Calcium	6.11	1.97	2.74	3.37	2.71	-0.66
Strontium	5.69	2.15	2.51	3.18	2.74	-0.44
Barium	5.21	2.22	2.43	2.78	2.48	-0.30
III						
Titanium	6.84	1.47	3.67	3.17	3.95	0.78
Zirconium	6.95	1.60	3.37	3.58	4.21	0.63
Niobium	6.77	1.46	3.70	3.07	4.01	0.94
Molybdenum	7.06	1.39	3.89	3.17	4.20	1.03
Tungsten	7.94	1.39	3.89	4.05	4.52	0.47
Rhenium	7.87	1.37	3.)4	3.93	5.10	1.17
Cobalt	7.84	1.25	4.32	3.52	4.40	0.88
Rhodium	7.7	1.34	4.03	3.70	4.80	1.10
Palladium	8.1	1.37	3.94	4.16	4.98	0.82
Platinum	8.96	1.39	3.89	5.07	5.34	0.27
Copper	7.72	1.28	4.22	3.50	4.46	0.96
Silver	7.58	1.44	3.75	3.83	4.73	0.90
IV						
Aluminum	5.97	1.43	3.78	2.19	4.08	1.89
Carbon	11.27	0.91	5.93	5.34	4.60	-0.74
Nickel	7.63	1.24	4.36	3.27	5.01	1.74
Gold	9.23	1.44	3.75	5.48	4.82	-0.66
Cadmium	8.99	1.54	3.51	5.48	4.07	-1.41
Mercury	10.44	1.57	3.44	7.00	4.53	-2.47
Bismuth	8.8	1.70	3.18	5.62	4.24	-1.38

a. See Periodic Chart of the Elements.[3]

b. Eq. 5.

c. Eq. 6.

d. See Handbook of Chemistry and Physics.[4]

a little less than 0.5 ev for the alkali metals.[5]

For the alkaline earths (II): The assumption that $\Delta\zeta = 0$ is probably not quite as good, but is still reasonable. The free atom has a moderately strong positive "surface double layer" that can be identified with the shielding coefficient of the valence electron left on the atom after the atom has been singly ionized. $\phi - \phi_0$ should be small or slightly negative.

Because of their complexity, no further comment is offered for the remaining elements, except to note that $\phi - \phi_0$ for elements from the center of the periodic table (part III of Table XIII-1) is of the order of magnitude of the surface double layer, 0.25-1 electron volt.[1,5,6]

This method of predicting monopole work will be explored further.

M. F. Koskinen

References

1. M. F. Koskinen, Work function of a conductor, Quarterly Progress Report No. 68, Research Laboratory of Electronics, M.I.T., January 15, 1963, pp. 108-113.

2. J. Bardeen, Phys. Rev. 58, 727 (1940).

3. Periodic Chart of the Elements, RCA Laboratories, Princeton, N.J. (1955).

4. Handbook of Chemistry and Physics (Chemical Rubber Publishing Company, Cleveland, 40th edition, 1958-1959).

5. J. Bardeen, Phys. Rev. 49, 655 (1936).

6. E. Wigner and J. Bardeen, Phys. Rev. 48, 85 (1935).

E. EMITTER WORK-FUNCTION PATCHES IN A CESIUM
   THERMIONIC CONVERTER

1. Introduction

An experimental program to obtain current-voltage characteristics from a metal cesium thermionic converter with a single-crystal (110) molybdenum emitter has been carried out. The results of this program are presented in this report. The output characteristics obtained experimentally are of the type that would be expected if a fraction of the emitter surface had a low work-function patch, the distribution of work function on this patch being given by an exponential Boltzmann type of distribution. The remainder of the surface has a discrete work function. The results of our analysis apply in the presence of transport effects in the interelectrode space. We find that the percentage of the area of the exponential patch increases as the cesium coverage increases.

2. Apparatus

A metal cesium thermionic converter has been designed and constructed specifically for the purpose of conducting laboratory tests. The diode configuration is of a parallel-plate nature with circular electrodes, 1 cm^2 in area, spaced 0.01 inch apart. A single-crystal molybdenum emitter was exposed to the cesium vapor. A complete description of the device has been given elsewhere.[1] The experimental apparatus, test circuit, and procedure have been presented in detail.[2] The testing consisted of setting a certain emitter and cesium temperature, and then obtaining the resulting I-V curve.

3. Theory

The purpose of this project was to offer an explanation for current-voltage curves of the form that display a linear increase in output current with decreasing output voltage in the region in which saturation current is expected, region B of Fig. XIII-10a. The theoretical treatment is presented in the author's thesis,[2] and only the major points are given here.

(a)   (b)   (c)
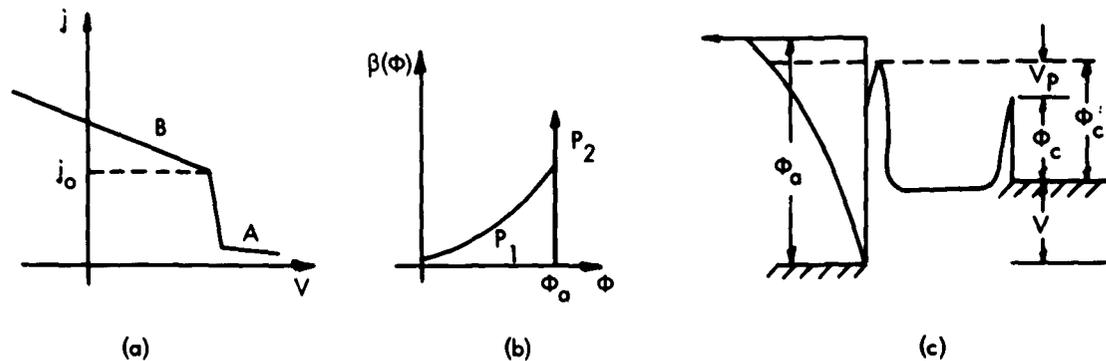
Fig. XIII-10. (a) Typical I-V curve; (b) Distribution of work function; (c) Motive diagram for operation in region B.

From physical considerations it is expected that a surface partially covered with cesium will have a Boltzmann distribution of work function over a fraction of its area. This is given by

$$B(\phi) = \frac{P_1}{kT} \exp\left[-\left(\frac{\phi_a - \phi}{kT}\right)\right],$$

where $B(\phi)$ is the distribution function, P_1 is the fractional area, and ϕ_a is the maximum work function. This is illustrated in Fig. XIII-10b. The effects of this distribution on

the output characteristics in region B can be seen by considering the motive diagram,
Fig. XIII-10c. As the output voltage is reduced, more of the emitter surface becomes
free of the space-charge barrier $\phi_c + V_p + V$, where V_p is the plasma drop. As more
of the surface is revealed, more saturation current is obtained from the low work-
function patches. The current-voltage relation in region B is given by

$$j_B = j_o \left[1 + \frac{P_1 (\phi_a - \phi_c' - V)}{kT} \right],$$ (1)

a linear relation between j and V. The fractional area of the exponential patch is given
by

$$P_1 = -\frac{kT}{j_o} \frac{dj_B}{dV}.$$ (2)

The area is a function of the cesium coverage. From Carabateas,[3] $(T/T_R) = f'(\theta)$
for low coverage. Thus θ is uniquely determined by the temperature ratio and, in fact,
decreases in an approximately linear fashion with increasing $\frac{T}{T_R}$. If P_1 is a function of
coverage, then its value should also decrease with increasing $\frac{T}{T_R}$.
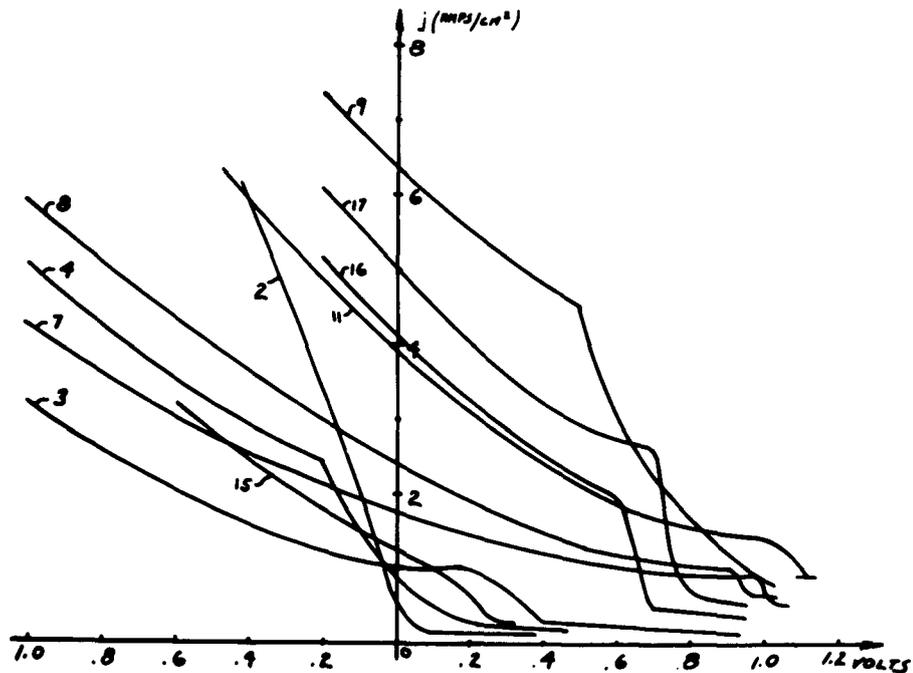
Fig. XIII-11. Experimental I-V curves.

4. Experimental Results

I-V curves obtained at various values of T and T_R appear in Fig. XIII-11. Note that all of the curves display the linear relation in region B given by Eq. 1. Calculated values for P_1 from Eq. 2 appear in Table XIII-2.

The graph of Fig. XIII-12 is a plot of P_1 vs T/T_R, which displays the relation between P_1 and θ.
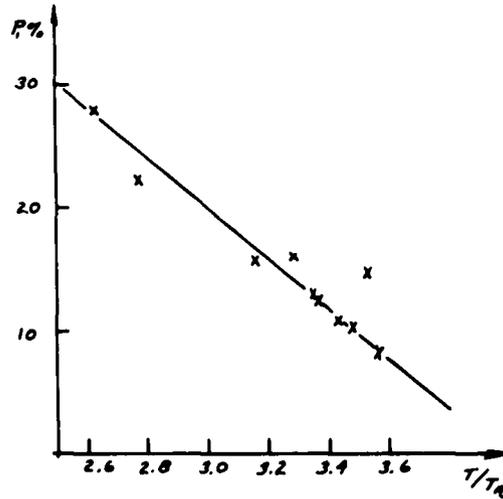
Fig. XIII-12. P_1 versus T/T_R.

Table XIII-2. Diode operating conditions and resulting P_1 from Eq. 2.

Curve	$\dfrac{T}{(°K)}$	$\dfrac{T_R}{(°K)}$	$\dfrac{P_1}{(\%)}$	T/T_R
2	1553	593	28	2.63
3	1693	495	11	3.43
4	1573	478	16.2	3.29
7	1803	510	7.92	3.54
8	1847	530	10.3	3.48
9	1858	555	13	3.35
11	1903	540	14.7	3.53
15	1600	578	22.3	2.77
16	1700	538	15.9	3.16
17	1818	543	12.6	3.35

5. Conclusion

Experimentally obtained current-voltage characteristics of the form shown in Fig. XIII-10 have been interpreted in terms of patches of low work function on the emitter surface. It has been shown that an exponential distribution of work function over a fractional area P_1 of the emitter surface will result in the I-V curves that were obtained, even in the presence of transport effects.

The patches have been shown to be a function of cesium coverage, as shown in Fig. XIII-12. The fractional area of the patches decreases linearly with decreasing cesium coverage. Thus we can conclude that the exponentially distributed patch is caused by thermal excitation of the adsorbed layer. A complete description of this experiment is given in the author's thesis.

J. W. Gadzuk

References

1. I. Psarouthakis, The Design, Construction, and Testing of a Thermionic Energy Converter — Single Crystal Molybdenum Emitter, S. M. Thesis, Department of Mechanical Engineering, M. I. T., May 1962.

2. J. W. Gadzuk, The Effect of Emitter Work Function Patches on the Performance of Cesium Thermionic Converters, S. B. Thesis, Department of Mechanical Engineering, M. I. T., May 1963.

3. E. N. Carabateas, Thermodynamics of surface films, J. Appl. Phys. 33, 2698-2702 (1962).

F. ELECTRON DENSITY IN WET POTASSIUM VAPOR

The energy required to remove one electron from a conducting drop lies somewhere between the ionization potential of a single atom of the metal and the work function associated with an infinite plane surface of the material. This has given rise to speculations regarding the use of condensation drops in alkali-metal vapor magnetohydrodynamic systems.

A number of papers have appeared on the influence of submicroscopic carbon particles on the electron densities in carbon-rich flames. Einbinder,[1] as many others, assumes the following expression for the energy to remove the zth electron from a conducting drop of radius R

$$W = \phi_s + \frac{ze^2}{4\pi\epsilon_o R},$$

where ϕ_s is the flat-surface work function. This expression for $z = 1$ yields excessive values for W when $R < 100$ Å. We have improved the expression by studying the image

forces induced in a conducting sphere by an emitted charge.
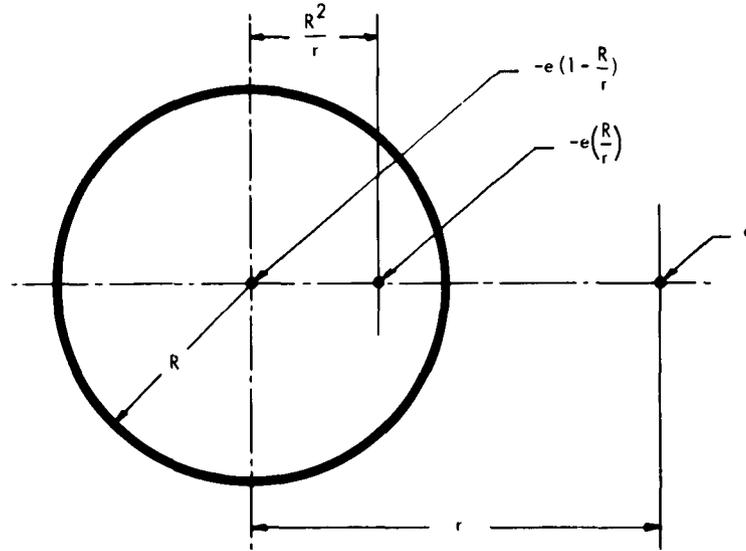
Fig. XIII-13. Induced charges in the spherical conductor.

The ideal force experienced by an electron as a result of induced charges in a drop (Fig. XIII-13) is

$$F(R, r) = \frac{e^2}{4\pi\epsilon_0}\left[\frac{(1-R/r)}{r^2} + \frac{R/r}{(r-R^2/r)^2}\right].$$

Close to the surface of the drop (< 5 Å), the ideal force expression breaks down. We assume that the force behavior at such close distances is independent of drop radius and consequently can remove this unknown behavior by comparing two spheres of different radii.

$$\Delta W = \int_{R+\delta}^{\infty} F(R, r) \, dr - \int_{R'+\delta}^{\infty} F(R', r) \, dr$$

In the limit $R' \to \infty$ and $\delta \to 0$, we find that

$$\Delta W = \phi - \phi_s = \frac{3}{8}\frac{e^2}{4\pi\epsilon_0 R},$$

where ϕ is the energy required to remove the first electron from a drop. The work required to remove the zth electron from a drop takes the form

$$W_z = \phi_s + \left[\frac{3}{8} + (z-1)\right]\frac{e^2}{4\pi\epsilon_0 R}.$$

With this result, equilibrium multiple ionization in a wet potassium vapor has been studied by using a system of Saha equations

$$\frac{N_e N_y^z}{N_y^{z-1}} = 2G\left(\frac{2\pi m_e kT}{h^2}\right)^{3/2} \exp\left(\frac{-W_z}{kT}\right)$$

and the requirements

$$\sum_z N_y^z = N_y$$

$$\sum_y \sum_z Z \cdot N_y^z = N_e.$$
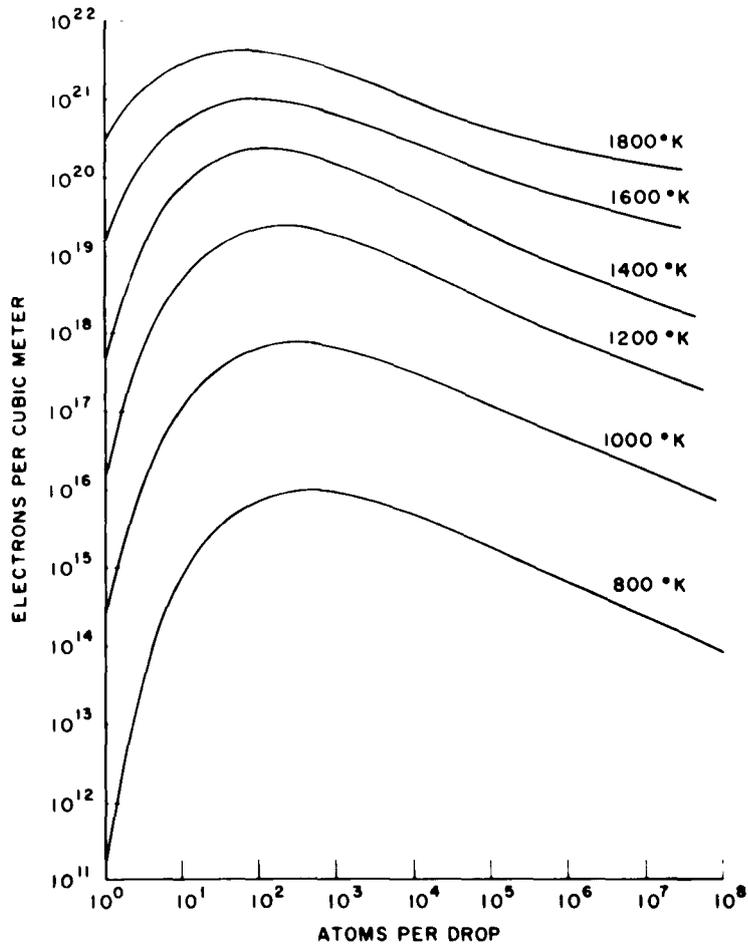
Fig. XIII-14. Electron density in wet potassium vapor.

Here, the lower subscript on particle density N selects particles of a definite size and the superscript further selects those with the same charge. G is the statistical weight equal to 1/2 for potassium atoms, and to unity for drops that differ only in electrostatic charge.

The electron density in a 10 per cent wet potassium vapor was found by using the facilities of the Computation Center, M.I.T. We assumed that all of the drops were of identical size. Figure XIII-14 shows the variation of electron density with vapor temperature and drop size.

It appears that in the plasma studied for this report electron-neutral atom collisions determine the electric conductivity. If this is the case, then the neutral atom-to-electron density ratio governs the conductivity, and the maximum conductivity occurs when the drop size maximizes the electron density in Fig. XIII-14. At 1600°K, the maximum conductivity is 3.5×10^{-1} mho/m as compared with 8.4×10^{-3} mho/m for the dry gas. At lower temperatures the effects of drops are even more pronounced. The implications of these results to magnetohydrodynamic generators continue to be studied.

A. W. Rowe, J. L. Kerrebrock

References

1. H. Einbinder, Generalized equations for the ionization of solid particles, J. Chem. Phys. 26, 948 (1957).

COMMUNICATION SCIENCES AND ENGINEERING

XIV. STATISTICAL COMMUNICATION THEORY[*]

Prof. Y. W. Lee	R. F. Bauer	T. G. Kincaid
Prof. A. G. Bose	E. M. Bregstone	A. J. Kramer
Prof. D. J. Sakrison	J. D. Bruce	D. E. Nelsen
Prof. M. Schetzen	A. M. Bush	J. K. Omura
Prof. H. L. Van Trees, Jr.	J. K. Clemens	A. V. Oppenheim
V. R. Algazi	A. G. Gann	R. B. Parente
R. Alter	L. M. Goodman	J. S. Richters
D. S. Arnstein	C. E. Gray	W. S. Smith, Jr.
M. E. Austin		D. W. Steele

A. WORK COMPLETED

1. OPTIMUM SECOND-ORDER VOLTERRA SYSTEMS

This study has been completed by E. M. Bregstone. In May 1963, he submitted the results to the Department of Electrical Engineering, M. I. T., as a thesis in partial fulfillment of the requirements for the degree of Master of Science.

2. STUDY OF PHASE MODULATION WITH A GAUSSIAN SIGNAL

The present study has been completed by J. K. Clemens. It was submitted as a thesis in partial fulfillment of the requirements for the degree of Master of Science, Department of Electrical Engineering, M. I. T., May 1963.

M. Schetzen

3. MINIMIZATION OF ERROR PROBABILITY FOR A COAXIAL-CABLE PULSE-TRANSMISSION SYSTEM

This study has been completed by J. S. Richters. In May 1963, he submitted the results to the Department of Electrical Engineering, M. I. T., as a thesis in partial fulfillment of the requirements for the degree of Master of Science.

4. EXPERIMENTAL INVESTIGATION OF A SELF-OPTIMIZING FILTER

This study has been completed by W. S. Smith, Jr. He submitted the results to the Department of Naval Architecture, M. I. T., May 1963, as a thesis in partial fulfillment of the requirements for the degree of Master of Science and professional degree of Naval Engineer.

D. J. Sakrison

5. INFLUENCE OF NORMAL MODES OF A ROOM ON SOUND REPRODUCTION

This study has been completed by D. W. Steele. In May 1963, he submitted the results to the Department of Electrical Engineering, M. I. T., as a thesis in partial fulfillment of the requirements for the degree of Master of Science.

A. G. Bose

6. TRANSIENT BEHAVIOR OF A PSEUDO NOISE-TRACKING SYSTEM

This study has been completed by D. S. Arnstein. In May 1963, he submitted the results to the Department of Electrical Engineering, M. I. T., as a thesis in partial fulfillment of the requirements for the degree of Master of Science.

7. EXPERIMENTAL INVESTIGATION OF THRESHOLD BEHAVIOR IN PHASE-LOCKED LOOPS

This study has been completed by A. G. Gann and the results have been submitted to the Department of Electrical Engineering, M. I. T., May 1963, as a thesis in partial fulfillment of the requirements for the degree of Master of Science.

H. L. Van Trees, Jr.

B. A TWO-STATE MODULATION SYSTEM

A simple two-state modulation system was described in Quarterly Progress Reports No. 66 (pages 187-189) and No. 67 (pages 115-119). This report is a continuation of the analysis of this system and presents some of its applications.

1. Switching Frequency and Linearity

The block diagram of the modulation system is shown in Fig. XIV-1.

The general expressions for the time average $\bar{y}$ of the output $y(t)$ and the switching period were given in Quarterly Progress Report No. 67 (page 116). Those expressions form the basis from which the dc performance of the system can be predicted. The non-linearity of the system, defined below, is determined solely by the ratio of the zero-signal switching period to the time constant of the feedback network. Figure XIV-2 shows the percentage of nonlinearity as a function of the normalized output signal level with T_o/τ as the parameter. The percentage of nonlinearity is defined as

$$\frac{100 \cdot (\bar{y}-KE_s)}{\bar{y}}, \tag{1}$$

where K is the slope of $\bar{y}$ vs E_s as $E_s \to 0$. Notice that this definition of nonlinearity
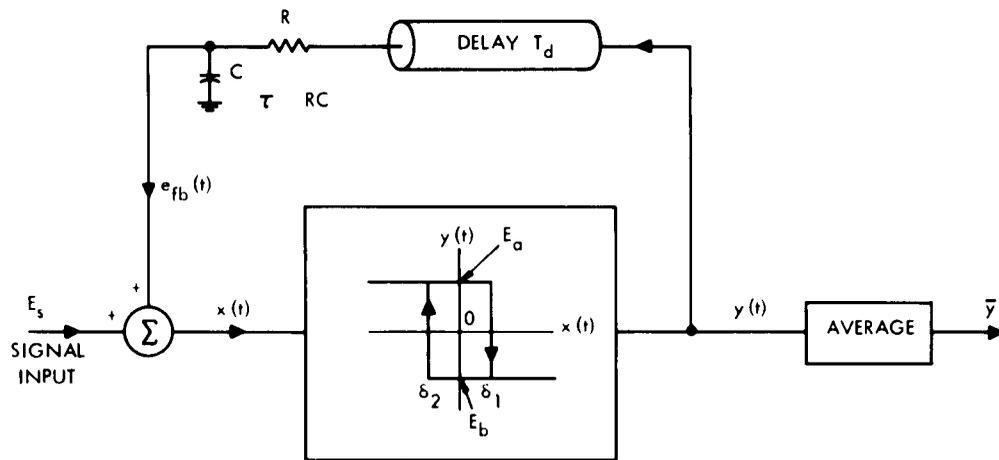
Fig. XIV-1. Modulation system with system parameters indicated.
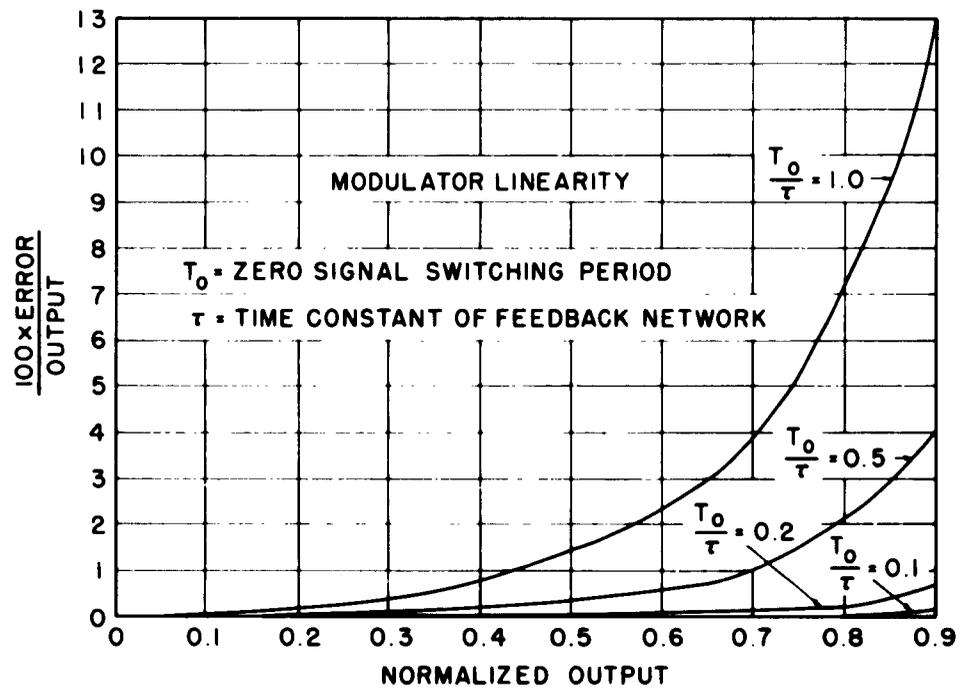
Fig. XIV-2. Modulator nonlinearity vs signal output.

is more stringent than the usual one in which the deviation is divided by the maximum output. The latter yields low distortion figures for systems that perform well with respect to large-amplitude signals but may badly distort small-amplitude signals.

For ac signals the time constant τ must be short enough so that the feedback network passes the highest signal frequency. If we arbitrarily set τ so that the half-power point of the feedback network occurs at twice the upper limit of the signal bandwidth B (cps), we have

$$T_o/\tau = 4\pi T_o B = 4\pi \frac{B}{f_o}, \tag{2}$$

where $f_o = \frac{1}{T_o}$ is the zero-signal switching frequency. Thus, for a given signal bandwidth B, this system achieves linearity at the cost of bandwidth.
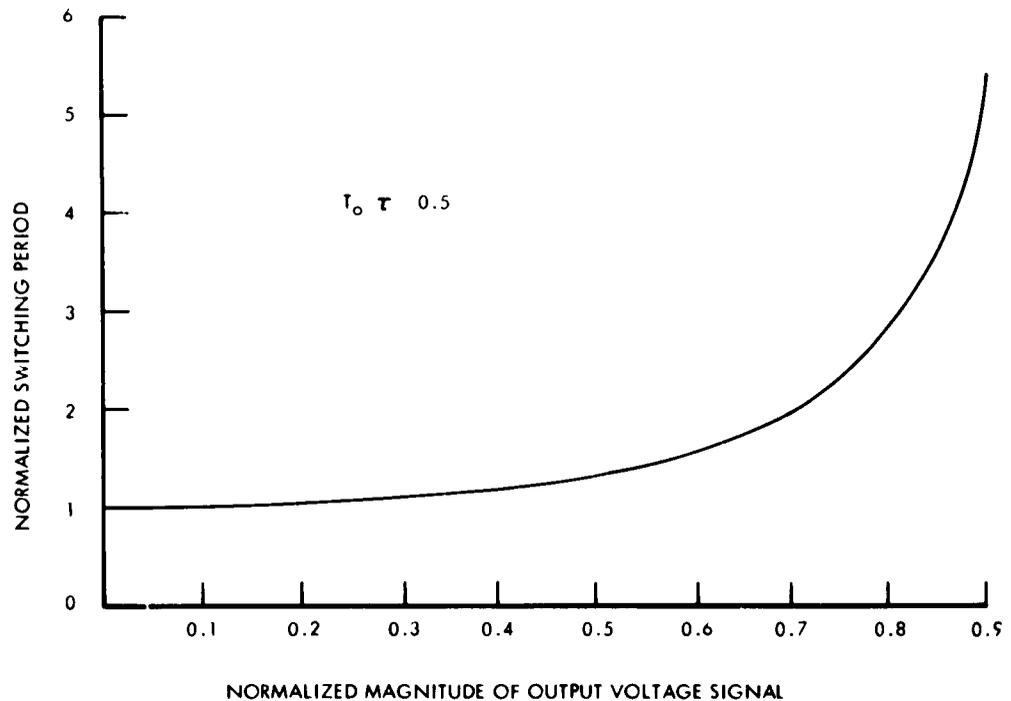
Fig. XIV-3. Switching period vs magnitude of output signal.

The switching frequency of the system increases monotonically with the magnitude of the output signal, as indicated in Fig. XIV-3, which shows the switching period as a function of the output-signal magnitude for dc signals. The curve is shown for $T_o/\tau = 0.5$.

2. The Modulator Used as a Regulator

The modulator can be used as a very simple regulator by taking advantage of the insensitivity of $\bar{y}$ to changes in the forward-loop saturation levels. Suppose that
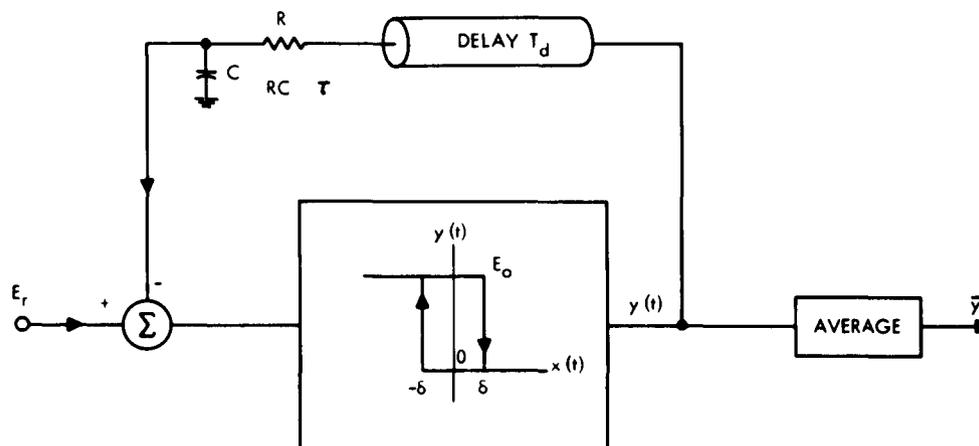
Fig. XIV-4. Regulator. Lower saturation level is set equal to zero to correspond to the voltage regulator example (see Fig. XIV-7).

we apply a dc reference signal E_r to the system shown in Fig. XIV-4. Analysis of this system for the output $\bar{y}$ yields the expression

$$\bar{y} = 1 - \frac{\ln\left(\dfrac{E_r - \delta}{E_o \exp(T_d/\tau) - E_o + E_r + \delta}\right)}{\ln\left(\dfrac{(E_r - \delta)(E_o - E_r + \delta)}{E_o \exp(T_d/\tau) - (E_o + E_r + \delta)(E_o \exp(T_d/\tau) - (E_r - \delta))}\right)}. \tag{3}$$

Preliminary calculations based on expression (3) and on measurements on an experimental model indicate that variations in the saturation level E_o can be suppressed more than 60 db at the output $\bar{y}$. They indicate that under all circumstances the regulation is improved by reducing the loop delay T_d. Curves showing the regulation as a function of the system parameters are being prepared.

3. Application to Nonlinear Control

The relay, or bang-bang servomechanism, is commonly used in control applications in which weight and power dissipation are significant considerations. The block diagram of such a controller is shown in Fig. XIV-5.
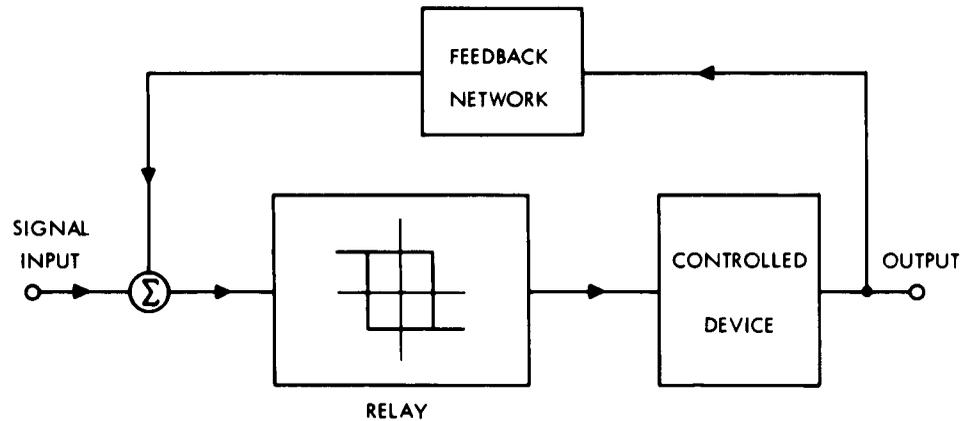
Fig. XIV-5. Simple relay servomechanism.

Briefly, the design of this system involves a compromise between static error, dynamic performance, and hunting. Hunting will always be present at the output unless the forward-loop nonlinearity contains a dead space. If a dead space is introduced, then sufficient damping, with its associated power loss, must be used to trap the system in the dead space. The damping requirement can be decreased by increasing the dead-space interval, but this is done by sacrificing the static error. These compromises are necessary because the system output (the output of the controlled device) is used to initiate the switching instants. Therefore switching must exist inherently at rates to which the controlled system responds.

By using the properties of the two-state modulator, it is possible to design a
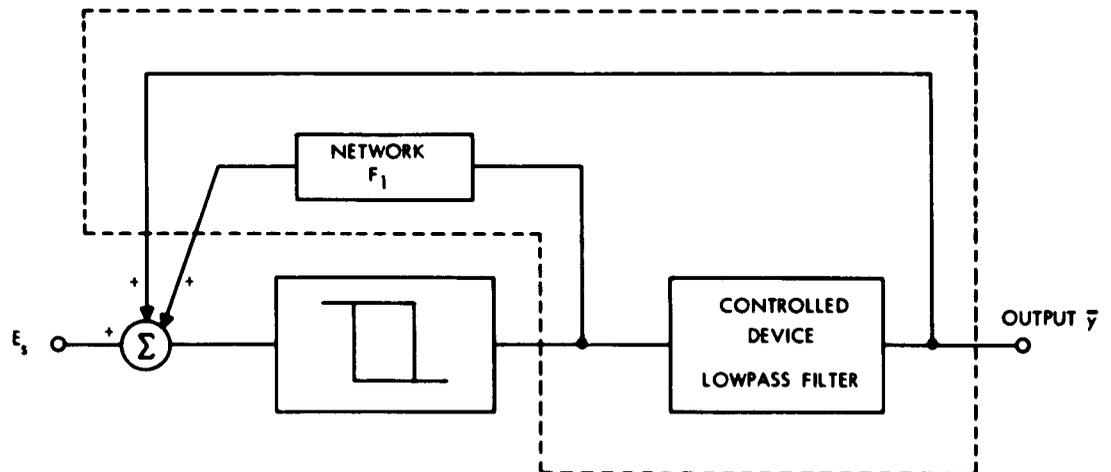
Fig. XIV-6. Nonlinear control system incorporating the modulator principle.

nonlinear controller that retains the advantages of low power dissipation, light weight, and, to a certain extent, the minimal settling time property, but eliminates the need for the dead space and damping usually employed to reduce hunting.

The block diagram of this controller is shown in Fig. XIV-6. The system within the dotted box comprises the total feedback network around the nonlinear forward loop. This network is divided into two paths. The result is that the network F_1 initiates an oscillation in the system at a frequency well beyond the response of the controlled device, and the over-all system still behaves linearly and has control with respect to load changes. The advantages of the low power dissipation of the relay servomechanism are preserved in the present system as long as the input to the controlled device is designed, as it can very easily be, to absorb no power at the high switching frequencies. In this design the ratio of the zero-signal switching frequency to the cutoff frequency of the controlled device can be made arbitrarily large without using active elements in the feedback loop and without introducing instability with respect to the system output. An analysis and the properties of this system will be presented in other reports. The application of the system to the problem of voltage control will be discussed here.

4. Voltage-Regulator Example

As an example of the nonlinear control discussed above, a simple voltage regulator was constructed (Fig. XIV-7). In this circuit the three transistors represent the nonlinear forward loop in Fig. XIV-6, L_ℓ and C_ℓ represent the controlled device, and R_1, R_2, C_i, C_2 represent the network F_1. This circuit was designed to supply 1 amp at
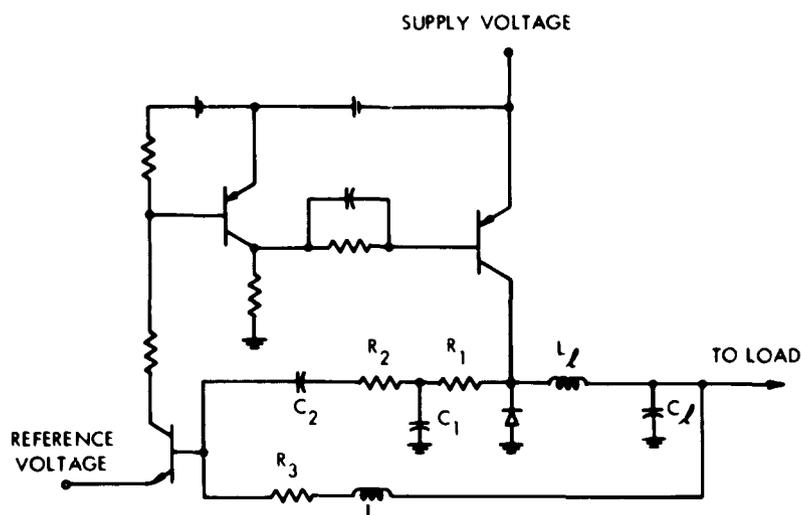
Fig. XIV-7. Voltage regulator that exemplifies nonlinear control.

15 volts to the load. The switching frequency was 100 kc. Measurements indicated better than 60-db attenuation of supply voltage variations at the load, and a change of only 0.14 per cent in the load voltage with a change by a factor of two in the load current.

5. D-C Transformer

Since each transistor in the realization of the modulator acts only as a switch, the system operates with high efficiency. From Eq. 3, it is possible to obtain any output voltage less than E - 6. Since the theoretical efficiency of the system is 100 per cent, we have the possibility of a very simple dc transformer, which works in the direction of stepping down voltages and stepping up currents, with the property that the power input equals the power output, and the additional feature that the output is regulated. Such a transformer was constructed by modifying the regulator of Fig. XIV-7 to obtain its reference voltage from its input as shown in Fig. XIV-8. This approach offers the possibility of building dc step-down transformers with efficiencies approaching those of conventional ac transformers, with the other features of lighter weight, smaller size, regulation, and adjustable "turns ratio." Thus in certain applications, in airborne equipment, for instance, there may be advantages in distributing dc instead of ac power. With ac power it is necessary, in each piece of equipment, first, to use a transformer, then to rectify, filter, and regulate. All of these operations could be carried out by a
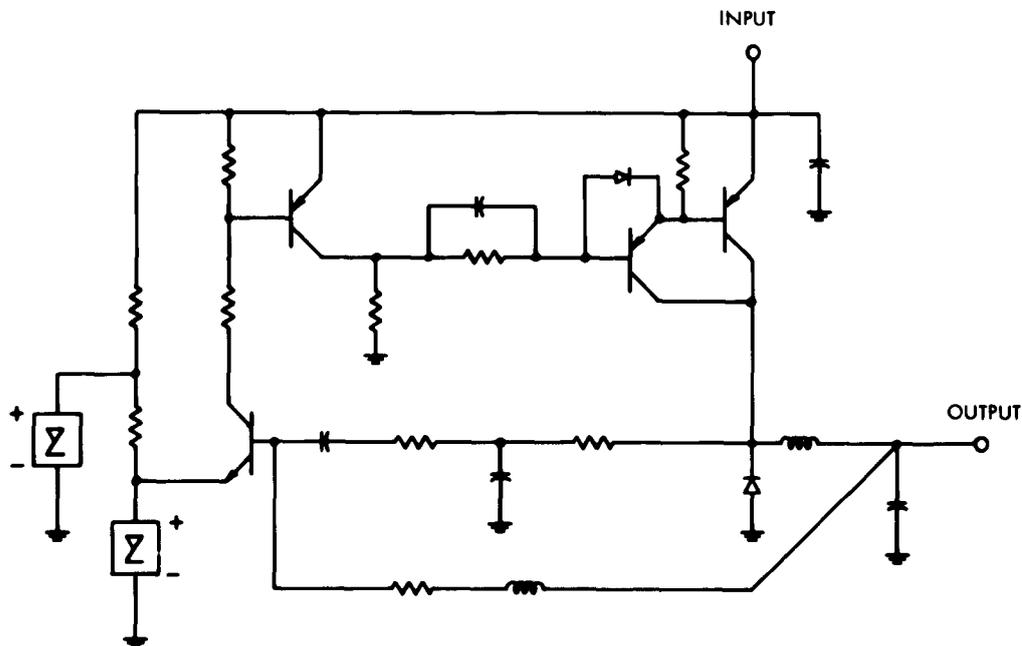
Fig. XIV-8. A dc transformer.

single dc transformer of the type discussed here. It should be noted, of course, that the dc transformer does not provide isolation and therefore would not suffice in an application for which it is essential. The transformer shown in Fig. XIV-8 was designed for only 15 watts power. However, it is possible, with existing transistors and similar circuitry, to construct transformers that deliver hundreds of watts.

6. Power Amplifier

The application of the modulation system to a dc-to-15-kc power amplifier is shown in Fig. XIV-9. The amplifier is designed to deliver 15 watts peak
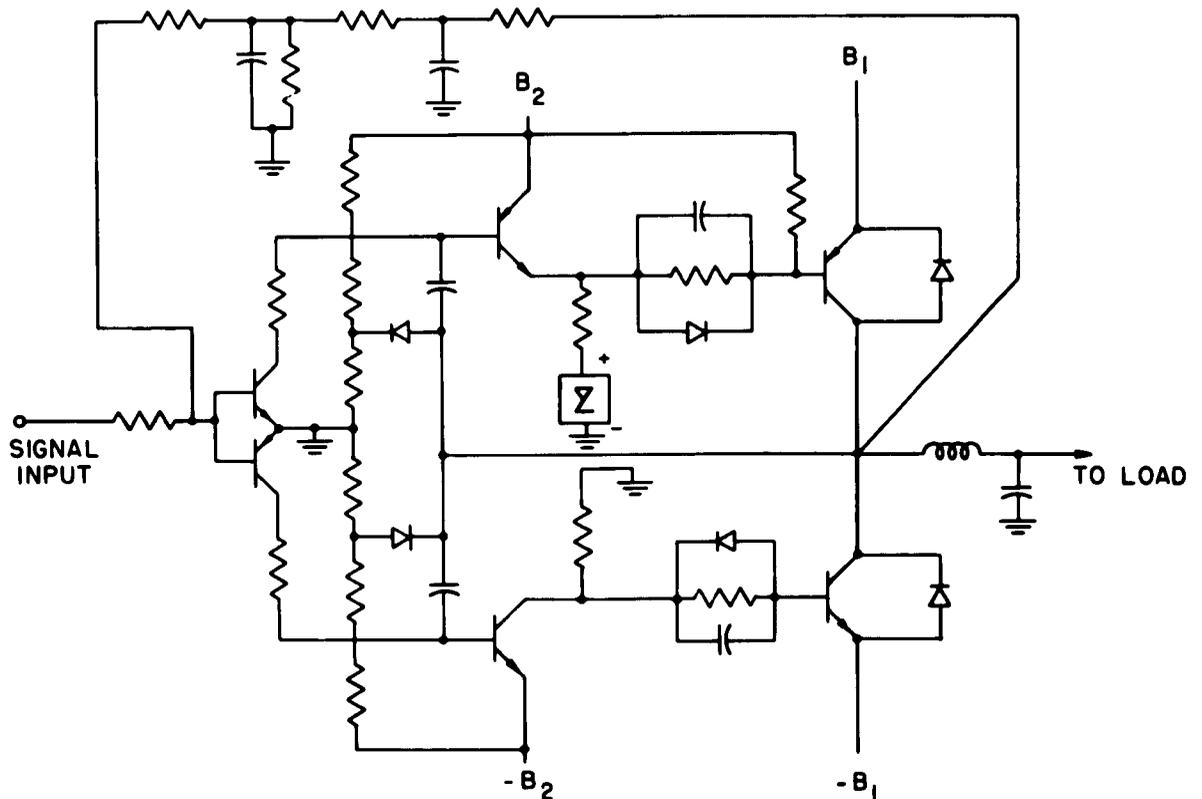
Fig. XIV-9. Circuit of 15-watt power amplifier.

power to a 16 Ω load. The mode structure and performance characteristics of this amplifier will be discussed in future reports.

A. G. Bose

C. OPTIMUM QUANTIZATION OF A SIGNAL CONTAMINATED BY NOISE

1. Formulation of the Problem

In a previous report[1] the problem of designing the optimum quantizer for a specific input signal was considered. An algorithm was developed by which the parameters defining the optimum quantizer can be calculated. The result is valid for a wide class of error criteria.

In many cases of interest, however, the signal is contaminated by noise before reaching the input of the quantizer. The noise may or may not be statistically independent of the signal. Mathematically, the quantizer input x may be written

$$x = s \oplus n, \tag{1}$$

where s is the signal, n is the noise, and $\oplus$ is a symbol that indicates some combination of the two variables, s and n. Two combinations of interest in communications
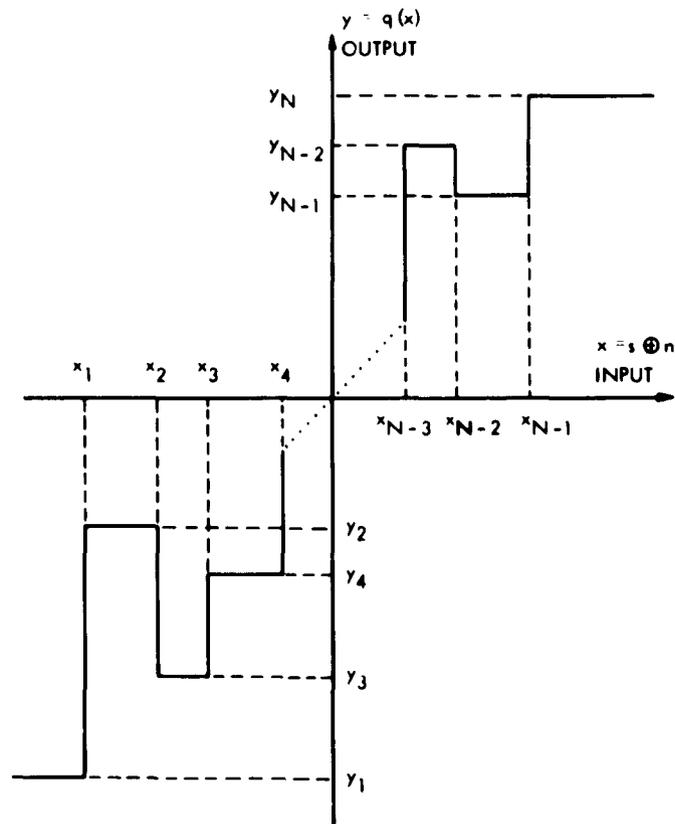
Fig. XIV-10. Input-Output relationship for a quantizer.

are (s+n) and (s·n). It will be seen from examination of Eq. 4 that any combination for which a joint probability density of x and s can be defined is an allowable combination.

The nonlinear, no-memory input-output characteristic of the quantizer, Fig. XIV-10, will be denoted by

$$y = q(x).$$

Then from (1) we have

$$y = q[s \oplus n]. \tag{2}$$

We desire to design a quantizer[2] that is such that its output y corresponds as closely as possible, with respect to some error criterion, to the signal s that is the desired quantizer output. Therefore, we shall take

$$s - q[s \oplus n] \tag{3}$$

to be the quantization error. The measure of the quantization error, that is, the quantity that we desire to minimize in order to minimize the quantization error, is taken to be the expected value of some function of the error $g[s-q(s \oplus n)]$, or equivalently, $g[s-q(x)]$. Then the measure of the error can be written

$$\mathscr{E} = \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} d\eta \left\{ g[\eta-q(\xi)] p_{x,s}(\xi,\eta) \right\}. \tag{4}$$

In Eq. 4, $p_{x,s}(\xi,\eta)$ is the joint probability density of the quantizer input x and the signal s.

Equation 4 can be simplified by considering the nature of the input-output characteristic of the quantizer. The output of the quantizer can take on only a fixed set of values; that is,

$$q(x) = y_i \qquad x_{i-1} \leq x < x_i, \qquad i = 1, 2, \ldots, N \tag{5}$$

where x_o is defined as X_ℓ, the lower bound of the input signal x. Likewise, x_N is the upper bound of x, X_u. (These bounds are not required to be finite.)

Substituting (5) in (4), we obtain the following equation for the "error" in terms of the quantizer representation values:

$$\mathscr{E} = \sum_{i=0}^{N-1} \int_{U_i} d\xi \int_{V_i} d\eta \left\{ g(\eta-y_{i+1}) p_{x,s}(\xi,\eta) \right\}. \tag{6}$$

U_i is that set of values of ξ for which the output assumes the value y_{i+1}. From Eq. 5 we see that U_i depends only on the quantizer characteristic and corresponds to the set of values of ξ which are such that $x_i \leq \xi < x_{i+1}$. Thus, Eq. 6 can be written

$$\mathcal{E} = \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} d\xi \int_{V_i} d\eta \left\{ g(\eta - y_{i+1}) P_{x,s}(\xi, \eta) \right\}. \tag{7}$$

Analogously, V_i is the set of values of η over which the output assumes the value y_{i+1}. Referring to U_i, we see that V_i is equivalently that set of values of η which are such that when the signal and the noise are combined they yield a quantizer input in the range $x_i \le \xi < x_{i+1}$. From Eq. 1 we see that V_i will depend upon the nature (or form) of the combination and upon the upper and lower bounds of the noise. Therefore, V_i cannot be explicitly expressed until the combination and the noise bounds are known. For purposes of formulating the method of determining the quantizer parameters, we shall utilize the "error" as given in Eq. 7.

Before we proceed with the task of determining the parameters of the optimum quantizer, let us demonstrate the method for determining the V_i. We shall consider two examples, assuming in each that Eq. 1 takes the specific form

$$x = s + n. \tag{8}$$

First, let us consider the case in which the noise is unbounded. Referring to Eq. 8, it is clear that any value of the signal can be transformed into the quantizer input region $x_i \le \xi < x_{i+1}$ and therefore can be represented at the output by y_{i+1}. Thus, the set V_i is $-\infty \le \eta < +\infty$. Then for this particular case, (7) becomes

$$\mathcal{E} = \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} d\xi \int_{-\infty}^{\infty} d\eta \left\{ g(\eta - y_{i+1}) P_{x,s}(\xi, \eta) \right\}. \tag{9}$$

As a second example, consider the case in which the noise is bounded with lower bound n_ℓ and upper bound n_u, that is, $n_\ell < n < n_u$. From Eq. 8 we see that values of the signal which are greater than $x_i - n_u$ can yield a quantizer input in the range $x_i < \xi$, since $n < n_u$. Also from (8) values of the signal which are less than $x_{i+1} - n_\ell$ can yield a quantizer input in the range $\xi < x_{i+1}$, since $n > n_\ell$. Thus, the set V_i which is the union of those values of the signal which correspond to $x_i < \xi$ and $\xi > x_{i+1}$ is expressed by the inequality

$$x_i - n_u < \eta < x_{i+1} - n_\ell.$$

Therefore, for this case (7) becomes

$$\mathcal{E} = \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} d\xi \int_{x_i - n_u}^{x_{i+1} - n_\ell} d\eta \left\{ g(\eta - y_{i+1}) P_{x,s}(\xi, \eta) \right\}. \tag{10}$$

2. Determination of the Optimum Quantizer

Equation 7 is an expression for the quantization error for an N-level quantizer. The optimum quantizer will be specified by specific values of x_i and y_k called X_i and Y_k, respectively. The X_i and Y_k specify the absolute minimum of (7) subject to the constraints

$$x_o < x_1 < x_2 \cdots < x_{N-1} < x_N. \tag{11}$$

These constraints are imposed only for organizational purposes.

At first, it might seem that the methods of calculus can be used to determine the X_i and Y_k. However, examination of the critical points of the error surface do not permit us to conclude that a critical point that is a relative minimum is also the absolute minimum within the region of variation specified by (11). Thus, in order to determine the X_i and Y_k we must turn to a more powerful technique, say, one that searches for the absolute minimum of the surface within the region of variation. Bellman's technique of dynamic programming[3,4] is such a technique.

In order to apply this technique, it is necessary to define three sets of functionals: the error functionals, $\{f_i(x_i)\}$; the transition-value decision functionals, $\{X_i(x)\}$; and the representation-value decision functionals, $\{Y_i(x)\}$. Each set of functionals has members for $i = 1, 2, \ldots, N$. These three sets of functionals are defined in the following manner:

$$\left.
\begin{aligned}
f_1(x_1) &= \min_{\substack{y_1 \\ X_\ell = x_o \leqslant x_1 \leqslant X_u}} \left\{ \int_{x_o}^{x_1} d\xi \int_{V_o} d\eta \, [g(\eta - y_1) P_{x,s}(\xi, \eta)] \right\} \\[2ex]
f_2(x_2) &= \min_{\substack{x_1, y_2 \\ X_\ell \leqslant x_1 \leqslant x_2 \leqslant X_u}} \left\{ f_1(x_1) + \int_{x_1}^{x_2} d\xi \int_{V_1} d\eta \, [g(\eta - y_2) P_{x,s}(\xi, \eta)] \right\} \\[1ex]
&\quad \vdots \\[1ex]
f_i(x_i) &= \min_{\substack{x_{i-1}, y_i \\ X_\ell \leqslant x_{i-1} \leqslant x_i \leqslant X_u}} \left\{ f_{i-1}(x_{i-1}) + \int_{x_{i-1}}^{x_i} d\xi \int_{V_{i-1}} d\eta \, [g(\eta - y_i) P_{x,s}(\xi, \eta)] \right\} \\[1ex]
&\quad \vdots \\[1ex]
f_N(x_N) &= \min_{\substack{x_{N-1}, y_N \\ X_\ell \leqslant x_{N-1} \leqslant x_N \leqslant X_u}} \left\{ f_{N-1}(x_{N-1}) + \int_{x_{N-1}}^{x_N} d\xi \int_{V_{N-1}} d\eta \, [g(\eta - y_N) P_{x,s}(\xi, \eta)] \right\}
\end{aligned}
\right\} \tag{12}$$

$$X_1(x) = X_\ell, \text{ a constant;}$$

$X_2(x)$ = the value of x_1 which corresponds to that in the
definition of the functional $f_2(x_2)$, $x_2 = x$;

.
.
.

$X_N(x)$ = the value of x_{N-1} which corresponds to that in
the definition of the functional $f_N(x_N)$, $x_N = x$ (13)

$Y_1(x)$ = the value of y_1 which corresponds to that in the
definition of the functional $f_1(x_1)$, $x_1 = x$;

$Y_2(x)$ = the value of y_2 which corresponds to that in the
definition of the functional $f_2(x_2)$, $x_2 = x$;

.
.
.

$Y_N(x)$ = the value of y_N which corresponds to that in the
definition of the functional $f_N(x_N)$, $x_N = x$ (14)

Consider these three sets of functionals for a few moments. The key to under-standing them is understanding the meaning of the separate members of the error func-tionals (12). The first member of (12) states that for a given region of $p_{x,s}(\xi, \eta)$ specified by the boundaries x_0 and x_1 (recall that V_0 is specified in terms of x_0 and x_1, and the nature of the noise contamination), we determine the y_1 that minimizes the integral

$$\int_{x_0}^{x_1} d\xi \int_{V_0} d\eta \, [g(\eta-y_1) p_{x,s}(\xi,\eta)]. \tag{15}$$

This y_1 is recorded as $Y_1(x)$, $x = x_1$ and the value of the integral (15) for this value of y_1 is recorded as $f_1(x_1)$. Thus, if we say that the region defined by the boundaries x_0 and x_1 is to be quantized, we know that the optimum representation value for this inter-val is $Y_1(x_1)$.

Now consider the second functional of (12). This functional states that we are con-sidering the quantization of the signal in the input interval $x_0 \leqslant x \leqslant x_2$, for a variable x_2, into two levels. In order to perform this operation in the optimum manner, we must minimize the quantity

$$\int_{x_o}^{x_1} d\xi \int_{V_o} d\eta \, [g(\eta - y_1) p_{x,s}(\xi, \eta)] + \int_{x_1}^{x_2} d\xi \int_{V_1} d\eta \, [g(\eta - y_2) p_{x,s}(\xi, \eta)] \qquad (16)$$

with respect to x_1, y_1, and y_2. However, the first of these two integrals when minimized with respect to y_1 (and it alone contains y_1) is simply the first error functional, $f_1(x_1)$. Then, for a specific x_2, we must determine the x_1 and the y_2 that minimize the function

$$f_1(x_1) + \int_{x_1}^{x_2} d\xi \int_{V_1} d\eta \, [g(\eta - y_2) p_{x,s}(\xi, \eta)]. \qquad (17)$$

The x_1 that minimizes (17) is recorded as $X_2(x)$, $x = x_2$; the y_2 that minimizes the expression is recorded as $Y_2(x)$, $x = x_2$. The value of the expression is recorded as $f_2(x_2)$. Therefore, if the region $x_o \leqslant x < x_2$ is to be quantized into two levels, we know from the decision functionals that the optimum transition value is specified by $X_1 = X_2(x_2)$ and that the two optimum representation values are given by $Y_2 = Y_2(x_2)$ and $Y_1 = Y_1(x_1)$.

Clearly, discussion of this type can be presented for each of the members of (12). However, instead of considering every member, let us skip to the last functional in (12). Here, we are given the input range $x_o \leqslant x < x_N$; a variable x_N is assumed. We want to quantize this range into N levels in the optimum manner. This requires that we minimize the quantity

$$\int_{x_o}^{x_1} d\xi \int_{V_o} d\eta \, [g(\eta - y_1) p_{x,s}(\xi, \eta)] + \int_{x_1}^{x_2} d\xi \int_{V_1} d\eta \, [g(\eta - y_2) p_{x,s}(\xi, \eta)]$$

$$+ \ldots + \int_{x_{N-1}}^{x_N} d\xi \int_{V_{N-1}} d\eta \, [g(\eta - y_N) p_{x,s}(\xi, \eta)] \qquad (18)$$

with respect to the parameters y_1, y_2, ..., y_N; x_1, x_2, ..., x_{N-1}. This task is not as difficult as it may seem. Note that the minimum of the first term with respect to y_1 as a function of x_1 is given by $f_1(x_1)$. This is the only term of (18) involving y_1. Thus (18) can be written alternately as the minimization of

$$f_1(x_1) + \int_{x_1}^{x_2} d\xi \int_{V_1} d\eta \, [g(\eta - y_2) p_{x,s}(\xi, \eta)] + \int_{x_2}^{x_3} d\xi \int_{V_2} d\eta \, [g(\eta - y_3) p_{x,s}(\xi, \eta)]$$

$$+ \ldots + \int_{x_{N-1}}^{x_N} d\xi \int_{V_{N-1}} d\eta \, [g(\eta - y_N) p_{x,s}(\xi, \eta)] \qquad (19)$$

with respect to $y_2, y_3, \ldots, y_N; x_1, x_2, \ldots, x_{N-1}$. But the minimization of the first two terms of (19) with respect to y_2 and x_1 as a function of x_2 is given by $f_2(x_2)$. Again, these are the only terms involving y_2 and x_1. Thus we can again reduce the complexity of the expression to be minimized. Equation 19 can be equivalently written as the minimization of

$$f_2(x_2) + \int_{x_2}^{x_3} d\xi \int_{V_2} d\eta \, [g(\eta - y_2) p_{x,s}(\xi, \eta)] + \int_{x_3}^{x_4} d\xi \int_{V_3} d\eta \, [g(\eta - y_3) p_{x,s}(\xi, \eta)]$$

$$+ \ldots + \int_{x_{N-1}}^{x_N} d\xi \int_{V_{N-1}} d\eta \, [g(\eta - y_N) p_{x,s}(\xi, \eta)] \tag{20}$$

with respect to $y_3, y_4, \ldots, y_N; x_2, x_3, \ldots, x_{N-1}$.

This process is easily continued until we obtain as an equivalent for (20) the minimization of

$$f_{N-1}(x_{N-1}) + \int_{x_{N-1}}^{x_N} d\xi \int_{V_{N-1}} d\eta \, [g(\eta - y_N) p_{x,s}(\xi, \eta)] \tag{21}$$

with respect to x_{N-1} and y_N. For a specific x_N, the x_{N-1} and y_N that minimize (21) are recorded as $X_N(x)$ and $Y_N(x)$, respectively, $x = x_N$. The value of (21) for a specific x_N is recorded as $f_N(x_N)$.

Observe, now, that when $x_N = X_u$, we are considering the entire input signal range. Thus, $f_N(X_u)$ is the total quantization error for the optimum N-level quantizer. Then from the definition of $X_N(x)$, the $(N-1)^{th}$ transition value is

$$X_{N-1} = X_N(X_u).$$

Likewise, from the definition of $Y_N(x)$ the $(N)^{th}$ representation value is

$$Y_N = Y_N(X_u).$$

Continuing from our definition of $X_{N-2}(x)$ and $Y_{N-2}(x)$, we find that the next transition value is

$$X_{N-2} = X_{N-1}(X_{N-1})$$

and the next representation value is

$$Y_{N-1} = Y_{N-1}(X_{N-1}).$$

This process can clearly be continued until finally we have

$$Y_1 = Y_1(X_1),$$

which is the last parameter needed to completely define the optimum quantizer.

At present, our research is concentrated on the properties of the error surface. It is hoped that a knowledge of the error surface properties will simplify the determination of the parameters defining the optimum quantizer.

J. D. Bruce

References

1. J. D. Bruce, Optimum quantization for a general error criterion, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M. I. T., April 15, 1963, pp. 135-141.

2. L. I. Bluestein, A Hierarchy of Quantizers, Ph. D. Thesis, Columbia University, May 1962, has considered a similar problem. He has been able to obtain an exact solution in the case of a discrete signal but only an approximate solution in the case of a continuous signal. In each case his error criterion was mean-absolute.

3. R. Bellman and J. Dreyfus, Applied Dynamic Programming (Princeton University Press, Princeton, New Jersey, 1962).

4. R. Bellman and B. Kotkin, On the Approximation of Curves by Linear Segments Using Dynamic Programming – II, Memorandum RM-2978-PR, The Rand Corporation, Santa Monica, California, February 1962.

D. MINIMIZATION OF ERROR PROBABILITY FOR A COAXIAL-CABLE PULSE-TRANSMISSION SYSTEM

We shall consider a particular pulse code modulation system visualized as one repeater link in a high-speed data-transmission facility. This system is illustrated in Fig. XIV-11. The criterion used in evaluating system performance is the probability of error at the system output, calculated by considering as degradations thermal noise and intersymbol interference from pulses that are no more than four time slots away in both the past and future. The system input is assumed to be a pulse train s_i $(t-kT)$, $k = 1, 2, 3, \ldots$. Each signal $s_i(t)$ is zero everywhere except on an interval $0 < t < T$. This set of possible input pulses is assumed to be binary, and the two possible signals are assumed to have equal energies and to be a priori equiprobable. Successive pulses are assumed to be statistically independent.

The transmission medium was chosen to be a 22-gauge paper-insulated two-wire line, 6000 feet long. This is the same type of line actually used in an experimental system at Bell Telephone Laboratories.[1] Also, to obtain solutions to some of the problems considered, it is necessary to use a rational expression for H(s); thus an approximation of the form $H(s) = \dfrac{A\omega_1\omega_2}{(s+\omega_1)(s+\omega_2)}$ was assumed, with $A = 0.4$, $\omega_1 = 5.03 \times 10^5$, and
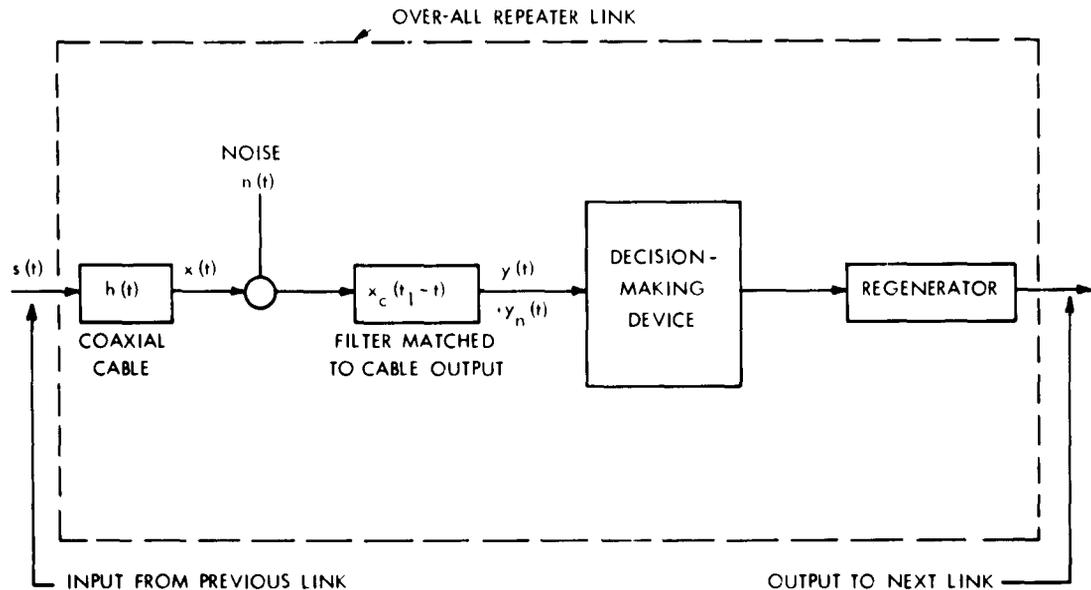
Fig. XIV-11. Block diagram of the system.

$\omega_2 = 5.03 \times 10^6$. This approximation was found to be "good" in both the time and frequency domains by Bell Telephone System engineers when the pulses were sent at a rate of 1.544 megabits per second. Therefore, the signal period T and the approximation to H(s) were used exactly as set forth by Mayo.[1]

At the cable output, additive noise, denoted n(t), was introduced. This is assumed to be white Gaussian thermal noise produced by the cable, with available noise power of $kT/2$ watts/cps (in a double-sided spectrum), where k is Boltzmann's constant, and $T = 293°K = 68°F.$[2] This thermal noise is assumed to be the only source of noise present.

The first stage of a repeater is usually some sort of filter to improve the signal-to-noise ratio going into the detector. In general, we are free to vary the shape of this filter at will, but, in order to cut down the number of variables involved in this problem, the filter was assumed to be matched to the cable output pulse x(t). In other words, the impulse response of the filter is $x_c(t_1-t)$, where

$$x_c(t) = \begin{cases} x(t) & 0 < t < t_1 \\ 0 & \text{elsewhere.} \end{cases}$$

We assume that t_1 is very large (in the sense that $x(t) \approx 0$ for $t \gg t_1$) and thus we may make the approximation $x_c(t_1-t) \approx x(t_1-t)$. Such a filter maximizes the signal-to-noise ratio at the output for any given x(t), and hence in the absence of interpulse interference

it minimizes the probability of error. This is not necessarily the best choice of filter, however, since other filters giving smaller output signal-to-noise ratios may give less intersymbol interference and thus smaller probability of error. The filter is followed by an instantaneous voltage sampler, a detector, and a regenerator.

Since the input signals are binary, equiprobable, and of equal energies, the signals will be of the form s(t) and -s(t) for a minimum error probability at the output.[3] The detector decision level should be set at zero, under the assumption that the two possible types of errors are equally harmful. For each problem considered, the optimization carried out was a variation of the input signal s(t) to determine the signal that resulted in the minimum error probability.

The first problem considered was to determine the input signal s(t) that maximizes the output signal-to-noise ratio (thereby minimizing error probability if only the noise is taken into account) for $\int_0^T s^2(t)\, dt$ constrained to be E. This problem is equivalent to that of maximizing the signal energy at the cable output with fixed input energy. Originally, the problem was considered to be the first step in a perturbation approach to our original problem, and its solution would minimize error probability when the effects of intersymbol interference are neglected. We hoped that this solution would not yield a prohibitive amount of intersymbol interference. This solution would then not differ by a great deal from the desired solution of the original problem, that of minimizing the probability of error. The desired solution then possibly could be obtained by perturbation techniques. This was not possible, since the solution of this problem resulted in an output pulse with such large amounts of interference that intersymbol interference from only two adjacent pulses was sufficient to cause an error in one of four cases.

The second problem was to determine the signal s(t) with fixed energy E which would maximize output signal-to-noise ratio, with the constraint that the amount of intersymbol interference one time slot away from the peak in either direction is a value K. Then the value of K was varied to obtain the minimum probability of error. By using simple variational procedures, an integral equation of the form

$$s(t) = \frac{1}{\lambda} \int_0^T s(u) \left[R_{hh}(t-u) + \mu (R_{hh}(t-u+T) + R_{hh}(t-u-T)) \right] du \tag{1}$$

was derived. In Eq. 1, λ and μ are Lagrangian multipliers that must be determined, and $R_{hh}(t) = \int_0^\infty h(a)\, h(a+t)\, da$. This equation is not always soluble, but in the special case in which H(s) is a rational function of s, the integral equation can be transformed into a simple differential equation that is easily solved.[4] The differential equation was derived and solved. For the two-pole approximation to H(s), the solution is of the form

$$s(t) = c_1 \left(e^{p_1 t} + c_2 e^{-p_1 t} + c_3 \cos p_2 t + c_4 \sin p_2 t \right) \tag{2}$$

for $0 \leqslant t \leqslant T$, and $s(t) = 0$ elsewhere. This is the general form of $s(t)$ for the maximization, and λ, μ, and the c_i were determined from the interference constraint, the fixed-input energy constraint, and the boundary conditions on the integral equation. These were determined numerically through use of the IBM 7090 digital computer at the Computation Center, M. I. T. Note that the first problem considered (no interference constraint) is a special case of this one, and follows from Eq. 1 if we set $\mu = 0$.

A third problem was that of constraining intersymbol interference to be zero at one and two time slots away from the pulse peak and again maximizing output signal-to-noise ratio. This constraint can be shown to constrain the interference to be zero for all time, since we have assumed that $H(s)$ is a two-pole rational function of s. Thus this problem is equivalent to one investigated by Holsinger.[5] Solutions can be obtained by his methods, which are much simpler than the integral-equation method that one would otherwise use. For most values of input energy, the solutions of this problem resulted in lower values of error probability than those found by constraining intersymbol interference one time slot away and varying the constraint for minimum error probability. For low input energies, the previous solutions gave better results, but this might be due to the fact that error probabilities were calculated by assuming interactions from only four pulses away, while in reality pulses farther away have nonnegligible interference. This will, of course, not affect error probabilities calculated for the pulse with zero interference, but error probabilities calculated for other pulses will be somewhat lower than their true values.

The final results were that

(1) Maximizing signal-to-noise ratio without considering intersymbol interference produces a solution that is unusable in an error probability sense.

(2) Maximizing signal-to-noise ratio while constraining intersymbol interference one time slot away and then varying the constraint for minimum error probability produces usable solutions, but in most cases not as good as those for result (3).

(3) Maximizing signal-to-noise ratio while constraining intersymbol interference to be zero one and two time slots away, which, for the assumptions made about the cable transfer function, results in zero interference everywhere.

These results are given in greater detail in the author's thesis.[6]

J. S. Richters

References

1. J. S. Mayo, A bipolar repeater for pulse code signals, Bell System Tech. J. **41**, 25-97 (1962).

2. A. Van der Zeil, Noise (Prentice-Hall, Inc., Englewood Cliffs, N. J., 1954).

3. G. L. Turin, An introduction to matched filters, Trans. IRE, Vol. IT-6, No. 3, pp. 311-329, June 1960.

4. W. B. Davenport and W. L. Root, Random Signals and Noise (McGraw-Hill Book Company, Inc., New York, 1958), pp. 371-382.

5. J. Holsinger, Some Problems Involved in Communication over a Class of Channels with Memory, Course 6.68 Special Problem Report, M. I. T., 1962.

6. J. S. Richters, Minimization of Error Probability for a Coaxial Cable Pulse Transmission System, S. M. Thesis, Department of Electrical Engineering, M. I. T., May 17, 1963.

E. EXPERIMENTAL INVESTIGATION OF A SELF-OPTIMIZING FILTER

1. Introduction

An experimental verification of the continuous adjustment procedure considered by Sakrison[1] is undertaken in this report. The adjustment procedure permits design of a filter system of the form shown in Fig. XIV-12 by the continuous adjustment of the k coefficients $x_1, x_2, \ldots, x_k$. The purpose of the adjustment procedure is to search out and converge to the coefficient setting that yields minimum average weighted error. The only assumption that we make on the weighting function is that it be convex. The adjustment procedure operates by estimating the gradient of the regression surface being searched and adjusting the values x_i, $i = 1, 2, \ldots, k$, to move the system error toward its minimum. Through the function c(t), a known plus and minus perturbation is introduced to the present setting of each of the x_i. When the errors resulting from these plus and minus perturbations of the ith parameter are measured, weighted by the chosen error weighting function, and subtracted, the resultant function, $Y_i(t)/c(t)$, is a random variable whose mean is a difference approximation of the ith component of the gradient in the direction of the optimum. The function a(t) is a monotonically decreasing function that is then used to weight the value of the gradient. The parameter change is then

$$x_i(t) = x_i(\phi) - \int_0^t \frac{a(\tau)}{c(\tau)} Y_i(\tau) \, d\tau.$$

It has been shown[1] for certain choices of the functions a(t) and c(t) and under our assumption of a convex error weighting function and under certain regularity assumptions on the processes, that the procedure will converge in the sense that

$$\lim_{t \to \infty} E\left\{(x_i(t)-\theta_1)^2\right\} = 0 \qquad i = 1, 2, \ldots k$$

in which θ_1 denotes the optimum setting for the coefficient x_i. The functions $a(t) = K_a/(at+b)^{oc}$ and $c(t) = K_c/(t+d)$ satisfy the conditions for convergence if

CONTINUOUS
BOUNDED
OPERATIONS
ON v (τ)
τ ≤ t
(LINEAR OR
NONLINEAR)

$v(t)$ $f_1(t)$ x_1 $f_2(t)$ x_2 $f_k(t)$ x_k Σ $q_{\underline{x}}(t)$ $d(t)$ $e(t)$
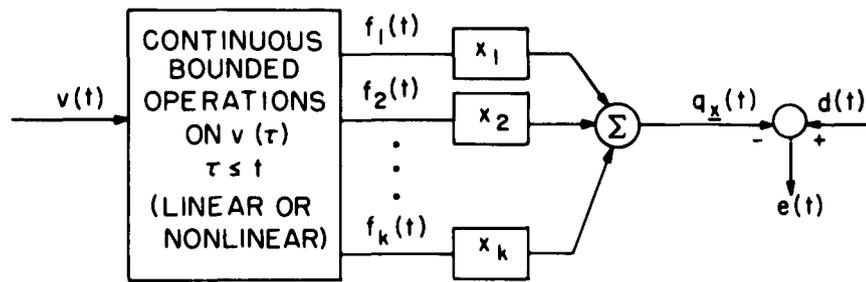
Fig. XIV-12. Form of the filter (or predictor or model) to be designed. (This is D. J. Sakrison's Fig. XIII-3, Quarterly Progress Report No. 66, p. 192.)

CONTINUOUS
BOUNDED
REALIZABLE
OPERATIONS
ON THE
INPUT

$v(t)$ $f_1(t)$ $x_1(t)$ $x_1 f_1$

$f_2(t)$ $x_2(t)$ $x_2 f_2$

$\frac{1}{s}$ Σ $q_{\underline{x}}(t)$ $d(t)$ $e(t)$

$+c(t)$ e_1^+ W $Y^1(t) = W(e_1^+)$

$-c(t)$ e_1^- W $Y^2(t) = W(e_1^-)$ $Y_1(t)$ $\frac{a(t)}{c(t)}$

$+c(t)$ e_2^+ W $Y^3(t) = W(e_2^+)$

$+c(t)$ e_2^- W $Y^4(t) = W(e_2^-)$ $Y_2(t)$ $\frac{a(t)}{c(t)}$

$$x_1(t) = x_1(t) - \int_1^t \frac{a}{c(t)} Y_1(\tau) \, d\tau$$
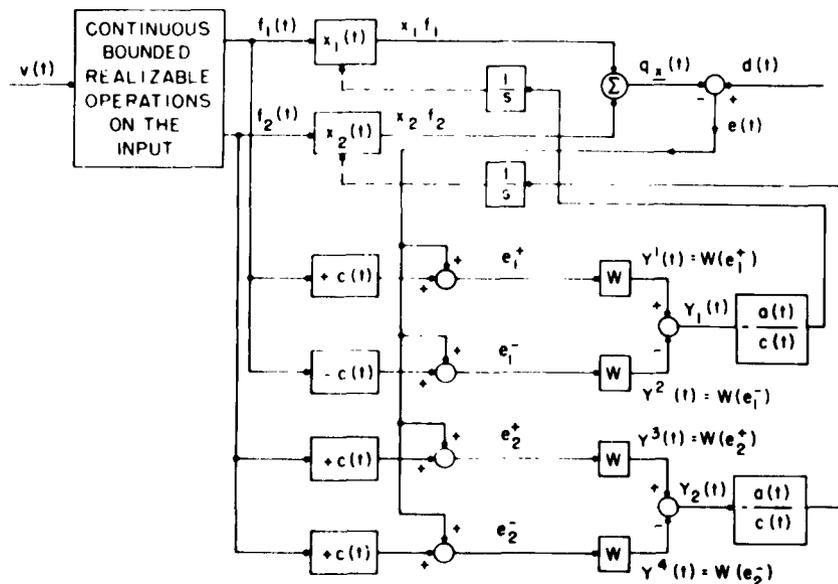
$$x_2(t) = x_2(t) - \int_1^t \frac{a(\tau)}{c(\tau)} Y_2(\tau) \, d\tau$$

$$e_1^+ = -f_1(t)\left[x_1(t) + c(t)\right] - f_2(t) \, x_2(t) + d(t)$$

$$e_1^- = -f_1(t)\left[x_1(t) - c(t)\right] - f_2(t) \, x_2(t) + d(t)$$

$$e_2^+ = -f_1(t) x_1(t) - f_2(t)\left[x_2(t) + c_2(t)\right] + d(t)$$

$$e_2^- = -f_1(t) x_1(t) - f_2(t)\left[x_2(t) - c_2(t)\right] + d(t)$$

Fig. XIV-13. Diagram of the continuous adjustment procedure for two parameters, x_1 and x_2. (This is D. J. Sakrison's Fig. XIII-4, Quarterly Progress Report No. 66, p. 192.)

$1/2 < o \cdot c \leqslant 1$ and $\delta > (1 - o \cdot c)/4$.

2. A Self-Optimizing Filter

Fant[2] describes the long-term average of speech as being characterized by a 12-db
octave decay over the audio spectrum. He also shows some curves of the actual spec-
trum which suggest that restricting the circuit usage to certain types of messages (that
is, a tactical voice circuit) would result in a power density spectrum that would follow
the 12-db curve on the average but would have significant plus and minus perturbations
from this mean. Such a desired output spectrum is approximately a staircase function
with steps of uneven height following the mean of 12 db per octave slope. Thus an inves-
tigation was made of a "self-optimizing" filter system for separating such an audio sig-
nal from broadband noise. This filter consisted of k third-order lowpass filters spaced
linearly throughout the useful portion of the audio band (250 to 2500 cps contains most of
the useful intelligibility). This system compares the desired output with that actually
being received and adjusts the coefficients of the k filters for minimum mean-square
error by the method just described. The mechanization of the adjustment procedure is
shown in block diagram form in Fig. XIV-13. Although the desired output signal was
used here in carrying out the adjustment procedure, it was done only for convenience:
for the case considered here, the adjustment procedure could be carried out without the
desired signal.[3]

3. Experimental Procedure

Experimental work was undertaken to ensure that the adjustment procedure would
lead to system parameter convergence on available analog equipment. In particular,
it was desired to know if the system parameters would converge within a time suitable
for the application proposed and if the adjustment procedure is critically dependent on
any of the adjustment parameters.

A filter system consisting of two lowpass filters combined with variable coefficients
x_1 and x_2 was studied in detail for a mean-square error criterion. This criterion was
chosen because it permitted an easy analytical check on the performance of the adjust-
ment procedure. On completion of these studies, checks were made for a system of
4 lowpass filters to ensure that the adjustment procedure still led to the proper results.

The desired output, $d(t)$, is obtained by passing the output of a white noise generator
(having power density spectrum S_0) through a filter of the form

$$\frac{K}{(j\omega + 628)^2}.$$

This gives the desired approximation of a mean speech spectrum. The input was formed
by adding to the desired output white noise from another generator having essentially a

a flat power density spectrum of magnitude N.

For the mean-square error criterion, the term $\frac{1}{c(\tau)} Y_1(\tau)$ is independent of $c(\tau)$ and we need only a(t). The function a(t) was generated by dividing a ramp of the form $A't+2$ into $Y_1(t)$ and multiplying the output of this operation by integrator gain D so that the form of a(t) was $D/(A't+2)$.

For purposes of this report, convergence time is defined as the time elapsed from activation of the system until the last time that the mean-square error (as found by time-averaging over a suitable interval) falls within 10 per cent of its minimum value. That region containing all those points having their mean-square error within 10 per cent of the minimum value is hereafter called the convergence area.
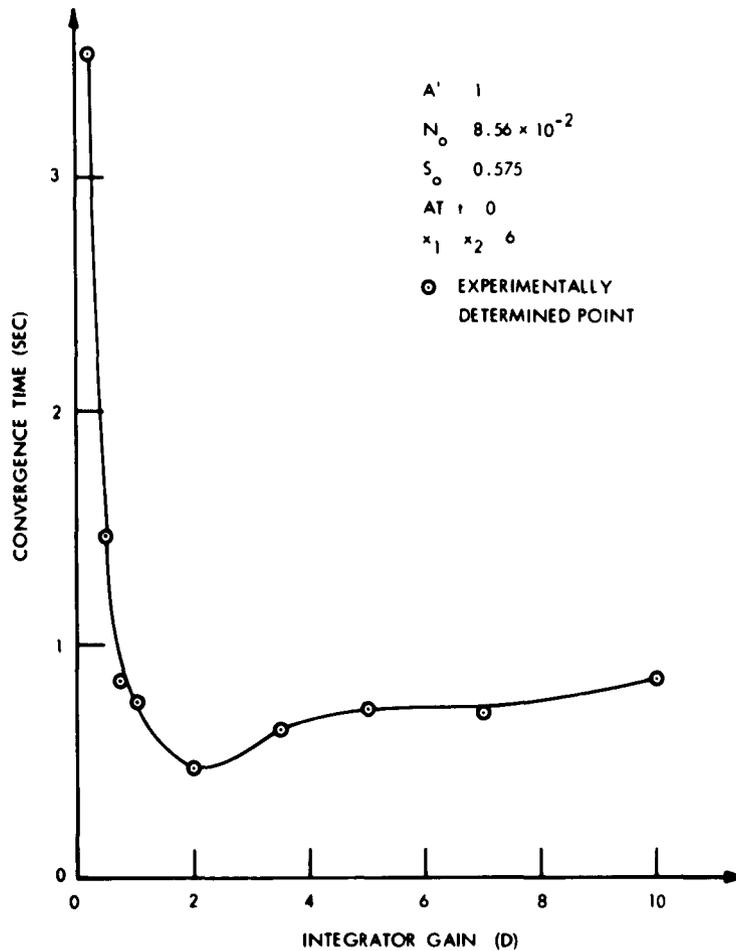
Fig. XIV-14. Convergence time versus integrator gain (D).

4. Results

Figure XIV-14 shows the convergence time versus the integrator gain, D. From this figure it is seen that large convergence time is obtained for integrator gains approaching zero. Settings between one and seven gave suitable convergence with initial oscillations about the convergence area, the amplitude of the oscillations increasing with increasing gains. Minimum convergence time was obtained for a setting of D just less than that value which first results in oscillations.

Figure XIV-15 shows the coefficient setting yielding minimum error, the convergence

INITIAL SETTING	CONVERGENCE TIME (SEC)
(6, -6)	4.37
(6, 0)	1.75
(6, 6)	0.45
(0, -6)	1.79
(0, 6)	21.9
(-6, -6)	38.5
(-6, 0)	8.30
(-6, 6)	29.2

S_o 0.575
N_o 8.56×10^{-2}
A' 1
D 2
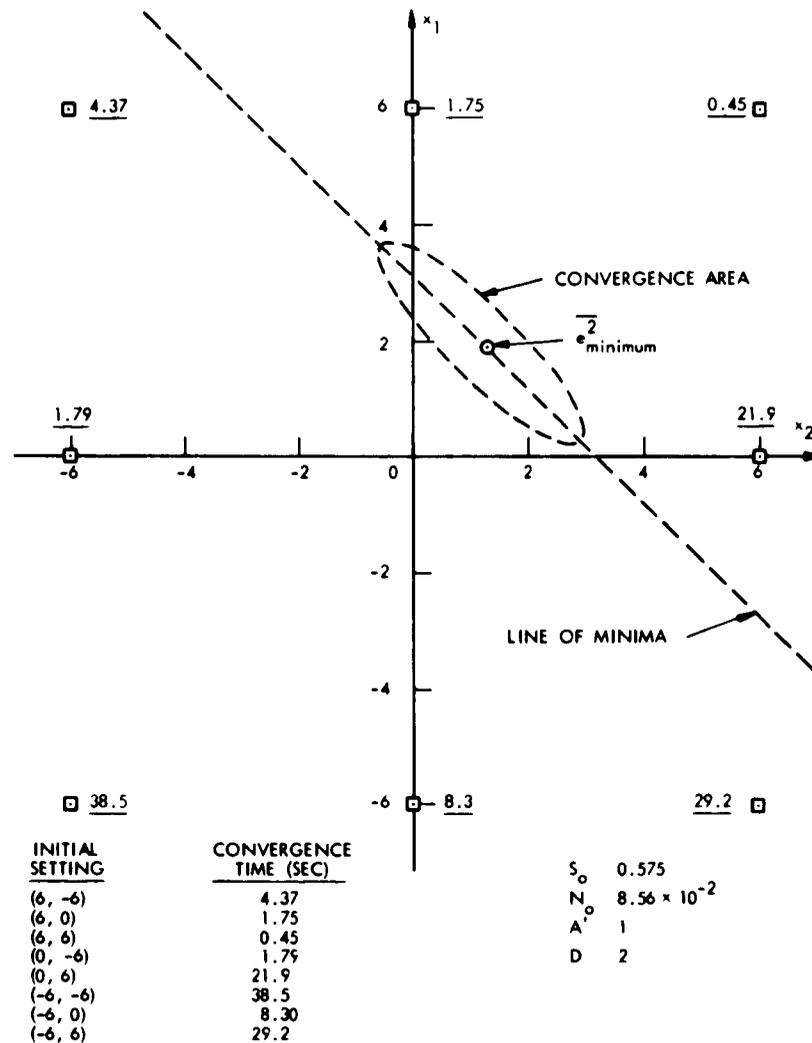
Fig. XIV-15. Convergence times from various initial settings of x_1 and x_2.

area, convergence times from various initial settings of the x_i, and the line of minima
(the major axis of the ellipsoids of constant mean-square error plotted against the x_i).
It is seen that the system takes very little time to reach the line of minima but takes
the majority of the convergence time to move down the valley along the line of minima.
For instance, if we look at the two points (-6,6) and (0,6), we see that the convergence
time is approximately the same from both points, but that a slightly longer time is taken
by the movement from the initial setting that is farthest from the line of minima. Looking
at the two points (6,0) and (6,-6), we see that the perpendicular distance to the line of
minima is approximately the same, but that for the initial setting farthest from the con-
vergence area the convergence time is approximately one-third longer compared with
the nearer setting.

A seeming deviation from this analysis is present if one considers the points (6,6)
and (-6,-6). There the discrepancy is due to an overshoot when (-6,-6) is the starting
point which does not occur when (6,6) is the starting point. (An example of the con-
vergence process with overshoot is shown in Fig. XIV-16.) The range in convergence
times from 0.45 second to 38.5 seconds is not indicative of the best possible perform-
ance of the system, since all of these runs were obtained with that value of D and A'
which gave minimum convergence time from (6,6).

Figure XIV-17 shows the minimum mean-square error, the experimentally meas-
ured mean-square error, and the mean-square error calculated for the experimentally
determined x_i, as a function of noise power with fixed desired output power for a 30-db
range of noise power. Throughout the range the experimental and theoretical results
agree quite closely, and thus indicate the capability of the adjustment procedure to
operate properly over this wide range.

It was found in the four-filter case that the error on completion of optimization is
approximately 40 per cent less than that obtainable with two filters. This is to be
expected, since with more filters we can more closely approximate the spectrum of the
desired signal.

5. Conclusions

We have shown that the adjustment procedure does produce the desired results for
the mean-square error criterion. The time taken to converge can be controlled by
proper setting of the adjustment procedure parameters. Since convergence times of
well under 30 seconds can be obtained with this system and the system does perform
satisfactorily over the required 30-db range in noise power density, it is applicable
for use as an adaptive filter system in a radio communications link in which the back-
ground noise varies widely but remains approximately constant for a quarter of an hour
or longer. Furthermore, it can be shown that the over-all gain in performance over
a fixed filter which can be achieved by even this simple example is of the order of 5 db
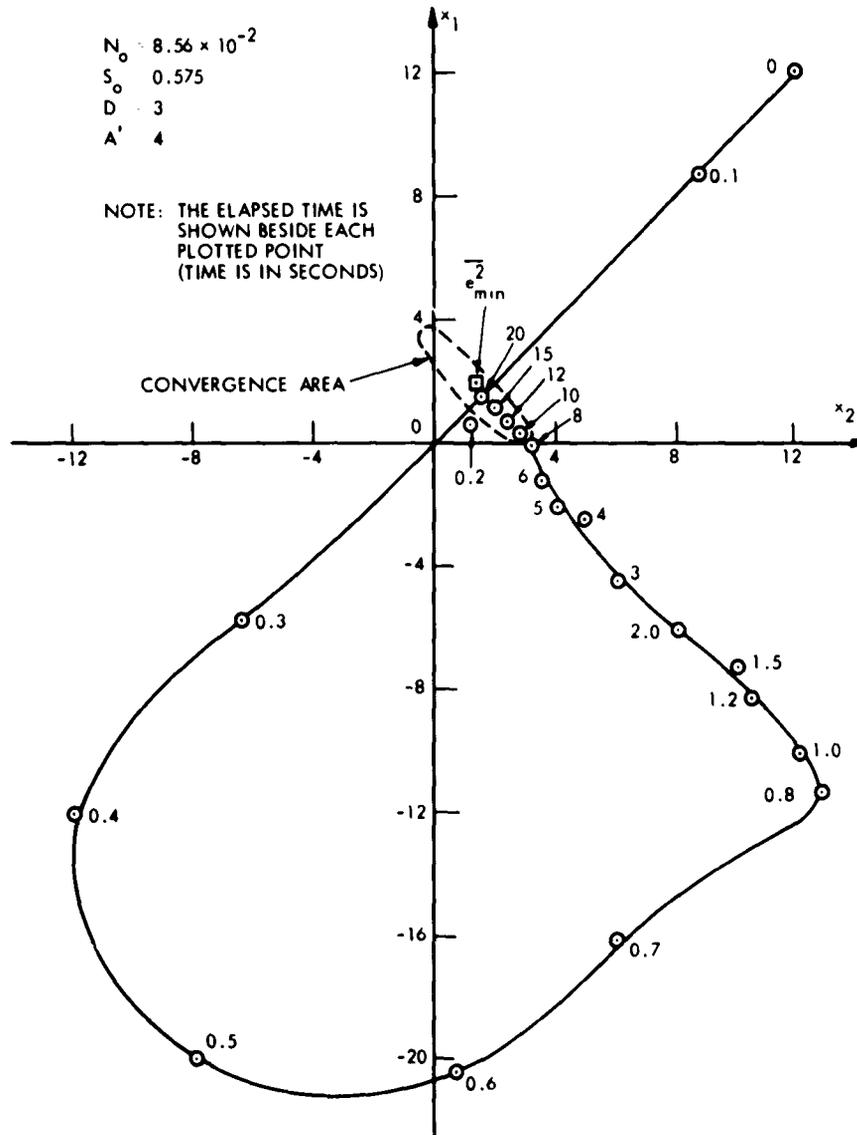
Fig. XIV-16. Typical convergence of the x_i (with overshoot).
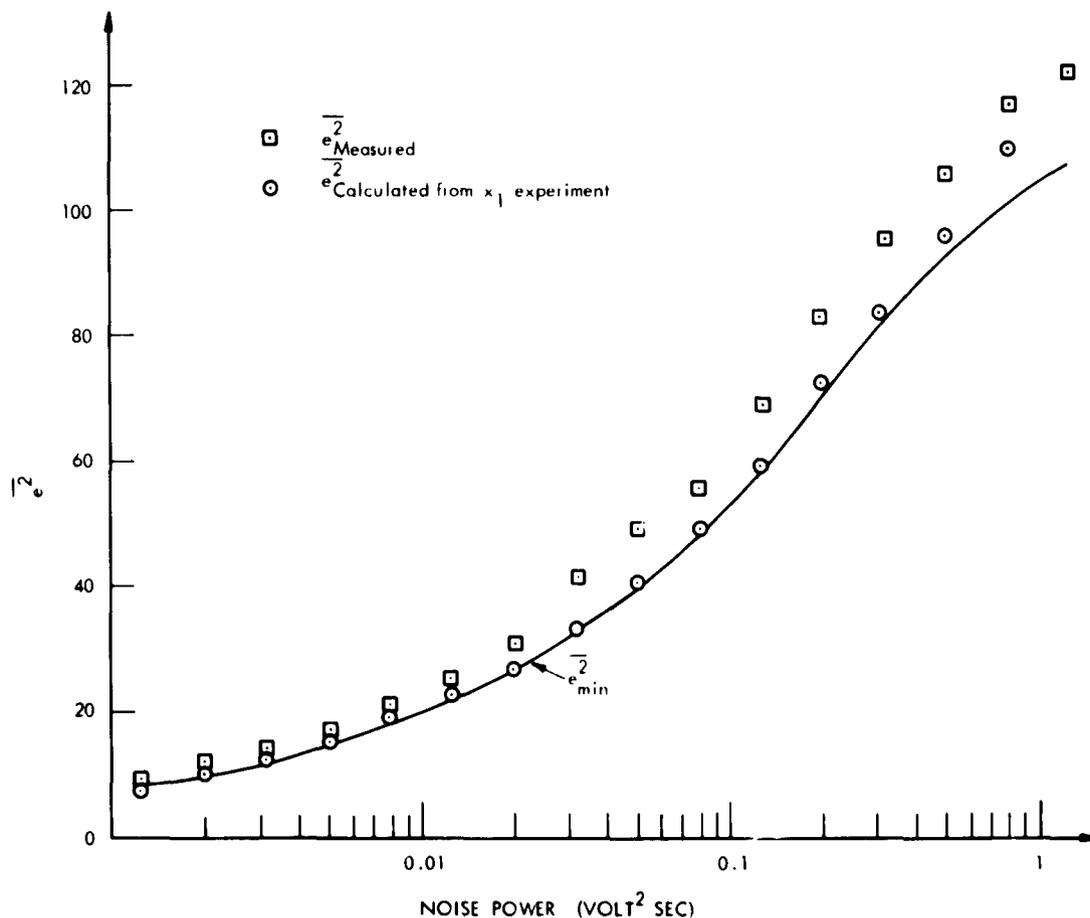
Fig. XIV-17. $\overline{e^2}_{measured}$, $\overline{e^2}_{minimum}$ and $\overline{e^2}_{calculated}$ from x_i experiment versus noise power for $S_o = 0.79$ volt^2 sec.

for a 14-db deviation of the noise power density from the value used for the fixed filter design.

A detailed description of these results has been given in the author's thesis.[4]

W. S. Smith, Jr.

References

1. D. J. Sakrison, Design of filters for non mean-square-error performance criteria by means of a continuous adjustment procedure, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M. I. T., July 15, 1962, pp. 189-201; Quarterly Progress Report No. 67, October 15, 1962, pp. 119-126.

2. G. Fant, Acoustic Theory of Speech Production (Mouton and Company, The Hague, 1961).

3. D. J. Sakrison, An iterative method for the design of filters for non-mean-square error criteria, IEEE Trans. on Information Theory (in press).

4. W. S. Smith, Jr., An Adaptive Filter System for Radio Communications, S. M. Thesis, Department of Naval Architecture, M. I. T., May 17, 1963.

F. STUDY OF PHASE MODULATION WITH A GAUSSIAN SIGNAL

1. Introduction

The problem of determining the power density spectrum of a randomly phase-modulated sinusoid has been of interest not only in information transmission but also in the study of the problem of many coupled oscillators. Wiener has studied this problem and has derived the spectrum in closed form for a modulating signal that is a Gaussian process.[1] He obtained this result by using the set of orthogonal functionals that he derived, the "G-functionals."

Since the publication of his work, no experimental testing of the theory has been carried out. Therefore from the results obtained by Schetzen[2,3] an experiment was prepared in which the power density spectrum of a phase-modulated wave was studied by using the Wiener derivations.

Wiener has shown[1] that it is possible to obtain a power density spectrum with a dip in it if the modulating signal is the white Gaussian noise response of a quadratic filter. It is shown here that it is possible to obtain a dip in the spectrum by using a linear filter.

2. Theoretical Determination of the Power Density Spectrum[4]

Figure XIV-18 shows the general phase-modulation system. The input to the system, x(t), is white Gaussian noise with zero mean and a power density spectrum $\Phi_{xx}(\omega) = 1/2\pi$. The system with impulse response h(t) is a linear time-invariant system, and the signal s(t) is therefore a Gaussian signal. The amplifier m is merely used to adjust the rms level of g(t). The oscillator output, $f_c(t)$, is phase-modulated by g(t).
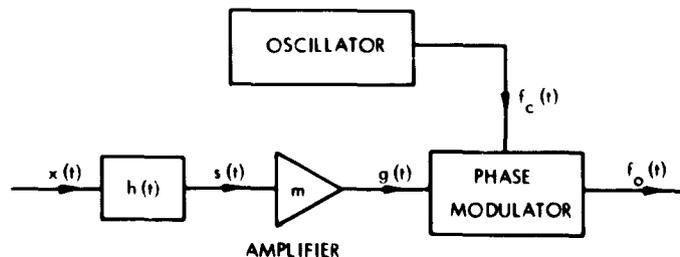
Fig. XIV-18. Phase-modulation system.

By the definition of phase modulation, if $f_c(t) = \cos \omega_c t$, then

$$f_o(t) = \cos (\omega_c t + g(t)). \tag{1}$$

We shall investigate the power density spectrum of $f_o(t)$. If we define a time signal $v(t) = e^{jg(t)}$, then $\psi_{oo}(\tau)$, the autocorrelation of $f_o(t)$, can be written in terms of $\psi_{vv}(\tau)$, the autocorrelation of $v(t)$,

$$\psi_{oo}(\tau) = \frac{1}{4} \left\{ e^{-j\omega_c \tau} \psi_{vv}(\tau) + e^{j\omega_c \tau} \psi_{vv}^*(\tau) \right\}. \tag{2}$$

Using the set of G-functionals and the techniques developed by Wiener,[1] we can put $\psi_{oo}(\tau)$ into a form that has an easily derived Fourier transform. The easiest way to do this is to express $g(t)$ in terms of $x(t)$. With respect to Fig. XIV-18

$$s(t) = \int_{-\infty}^{\infty} h(t-\sigma) x(\sigma) d\sigma. \tag{3}$$

Let us normalize $h(t)$ by letting

$$h(t) = k\phi(t) \tag{4}$$

where $k^2 = \int_{-\infty}^{\infty} h^2(t) dt$. By defining $b = mk$, $g(t)$ can be written

$$g(t) = b \int_{-\infty}^{\infty} \phi(t-\sigma) x(\sigma) d\sigma. \tag{5}$$

Wiener[1] has expanded $\psi_{vv}(\tau)$, the autocorrelation of $v(t)$, in terms of the G-functionals:

$$\psi_{vv}(\tau) = e^{-b^2} \sum_{n=0}^{\infty} \frac{b^{2n}}{n!} \left[\int_{-\infty}^{\infty} \phi(t) \phi(t+\tau) dt \right]^n. \tag{6}$$

Because of the special form of this expansion, it can be simplified. Defining

$$\psi_{\phi\phi}(\tau) = \int_{-\infty}^{\infty} \phi(t) \phi(t+\tau) dt, \tag{7}$$

we obtain

$$\psi_{vv}(\tau) = \exp \left\{ -b^2 [1 - \psi_{\phi\phi}(\tau)] \right\}. \tag{8}$$

Since $\psi_{vv}(\tau)$ is pure real, (2) can be written

$$\psi_{oo}(\tau) = \frac{1}{2} \psi_{vv}(\tau) \cos \omega_c \tau. \tag{9}$$

The spectrum of $\psi_{oo}(\tau)$ is merely the spectrum of $\frac{1}{4}\psi_{vv}(\tau)$ centered at the frequencies $\omega = \pm\omega_c$.

In order to simplify measurements, as well as calculations, it is desirable to set $\omega_c = 0$. This is merely a shift of the spectrum to a center frequency of $\omega = 0$. Notice that this forces the spectrum of $\psi_{oo}(\tau)$ to be symmetric. We then have

$$\psi_{oo}(\tau) = \frac{1}{2}\psi_{vv}(\tau). \tag{10}$$

The desired power density spectrum, $\Psi_{oo}(j\omega)$, is the Fourier transform of $\psi_{oo}(\tau)$. This can be written by expressing $\psi_{vv}(\tau)$ as a sum and interchanging the order of the sum and the integral,

$$\Psi_{oo}(j\omega) = \frac{1}{2}e^{-b^2}\sum_{n=0}^{\infty}\frac{b^{2n}}{n!}\frac{1}{2\pi}\int_{-\infty}^{\infty}\psi_{\phi\phi}^n(\tau)\,e^{-j\omega\tau}\,d\tau. \tag{11}$$

Since $\psi_{\phi\phi}(\tau)$ is an even function in τ, it can be represented as follows:

$$\psi_{\phi\phi}(\tau) = \begin{cases} \displaystyle\sum_{k=-\infty}^{\infty} A_k e^{s_k\tau} & \text{for } \tau \geq 0 \\[4mm] \displaystyle\sum_{k=-\infty}^{\infty} A_k e^{-s_k\tau} & \text{for } \tau < 0. \end{cases} \tag{12}$$

Defining $\psi_{\phi\phi_1}(\tau) = \psi_{\phi\phi}(\tau)$ for $\tau \geq 0$ and $\psi_{\phi\phi_2}(\tau) = \psi_{\phi\phi}(\tau)$ for $\tau < 0$ and both functions zero otherwise, we have

$$\psi_{\phi\phi}(\tau) = \psi_{\phi\phi_1}(\tau) + \psi_{\phi\phi_2}(\tau) \tag{13}$$

and

$$\psi_{\phi\phi}^n(\tau) = \psi_{\phi\phi_1}^n(\tau) + \psi_{\phi\phi_2}^n(\tau). \tag{14}$$

Defining $\Psi_1^{(n)}(s)$ and $\Psi_2^{(n)}(s)$ as the exponential transforms of $\psi_{\phi\phi_1}^n(\tau)$ and $\psi_{\phi\phi_2}^n(\tau)$, respectively, we can write

$$\Psi_1^{(n)}(s) = \sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}\cdots\sum_{k_n=-\infty}^{\infty}\frac{A_{k_1}A_{k_2}\cdots A_{k_n}}{\left(s-s_{k_1}-s_{k_2}-\cdots-s_{k_n}\right)}, \qquad \text{if Re }(s) > \sum_{i=1}^{n}\text{Re}\left(s_{k_i}\right) \tag{15}$$

and

$$\Psi_2^{(n)}(s) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \cdots \sum_{k_n=-\infty}^{\infty} \frac{(-1)A_{k_1} A_{k_2} \cdots A_{k_n}}{\left(s+s_{k_1}+s_{k_2}+\ldots+s_{k_n}\right)}, \qquad \text{if } \mathrm{Re}\,(s) < -\sum_{i=1}^{n} \mathrm{Re}\left(s_{k_i}\right).$$

(16)

Notice that the conditions on Re (s) are satisfied if the Fourier transform $\psi_{\phi\phi}(\tau)$ exists; this existence is assured by the Wiener theorem.[5]

In order to interpret these sums, it is convenient to relate them to an example. This example is the one that was used in the actual experiment and will be discussed in parallel with the rest of the derivation.

Let the only nonzero values for A_k in the expansion of $\psi_{\phi\phi}(\tau)$ be those for which $k = +1$ and $k = -1$. We then have

$$\psi_{\phi\phi_1}(\tau) = \begin{cases} A_1 e^{s_1 \tau} + A_{-1} e^{s_{-1}\tau}; & \tau \geq 0 \\ \\ 0; & \tau < 0. \end{cases}$$

(17)

From Eq. 15

$$\Psi_1^{(1)}(s) = \sum_{k_1=-1}^{1} \frac{A_{k_1}}{\left(s-s_{k_1}\right)} = \frac{A_1}{s-s_1} + \frac{A_{-1}}{s-s_{-1}}$$

(18)

$$\Psi_1^{(2)}(s) = \sum_{k_1=-1}^{1} \sum_{k_2=-1}^{1} \frac{A_{k_1} A_{k_2}}{\left(s-s_{k_1}-s_{k_2}\right)} = \frac{A_1 A_1}{s-2s_1} + \frac{2A_1 A_{-1}}{s-s_1-s_{-1}} + \frac{A_{-1}A_{-1}}{s-2s_{-1}}.$$

(19)

These two expressions, $\Psi_1^{(1)}(s)$ and $\Psi_1^{(2)}(s)$, are shown in Fig. XIV-19 in which the poles and residues (in parentheses) of each of the terms in the respective expressions are plotted in the s-plane. Notice that s_1 and s_{-1} are chosen to be a complex-conjugate pair, which they must be for a real autocorrelation function; this fact is represented by defining

$$s_1 = -\alpha + j\omega_o; \qquad s_{-1} = -\alpha - j\omega_o.$$

(20)

Notice that it is possible to consider $\Psi_1^{(2)}(s)$ the result of a two-dimensional convolution of $\Psi_1^{(1)}(s)$ with $\Psi_1^{(1)}(s)$ in the s-plane. The term two-dimensional convolution is meant to imply that the coordinates of each pole in the s-plane are found by a one-dimensional convolution along the respective coordinate.

The quantity $\Psi_1^{(3)}(s)$ is found to be

$$\Psi_1^{(3)}(s) = \frac{A_1^3}{(s-3s_1)} + \frac{3A_1^2 A_{-1}}{(s-2s_1-s_{-1})} + \frac{3A_1 A_{-1}^3}{(s-s_1-2s_{-1})} + \frac{A_{-1}^3}{(s-3s_{-1})},$$ (21)

which again can be recognized as the two-dimensional convolution of $\Psi_1^{(1)}(s)$ and $\Psi_1^{(2)}(s)$ in the s-plane.

In general, the expression $\Psi_1^{(n)}(s)$ is the result of a two-dimensional convolution of $\Psi_1^{(1)}(s)$ with $\Psi_1^{(n-1)}(s)$.
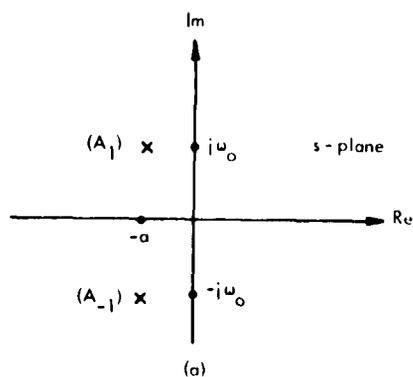
Fig. XIV-19. (a) $\Psi_1^{(1)}(s)$.

(b) $\Psi_1^{(2)}(s)$.
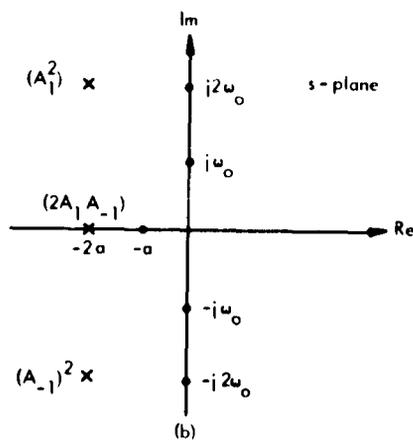
The representation of $\Psi_2^{(n)}(s)$ in the s-plane can be derived from $\Psi_1^{(n)}(s)$ by noticing that each of the poles appears at a point in the s-plane that is symmetric about $s = 0$ and the residue of the pole is the negative of the residue of the respective pole of $\Psi_1^{(n)}(s)$, as can be seen from Eqs. 15 and 16.

Although the example shown has only two nonzero A_k, it is easily seen that the technique, including the two-dimensional convolution, is valid in the general case.

Rewriting Eq. 11 in terms of the exponential transform, we have

$$\Psi_{oo}(s) = \frac{1}{2} e^{-b^2} \sum_{n=0}^{\infty} \frac{b^{2n}}{n!} \int_{-\infty}^{\infty} \left[\psi_{\phi\phi_1}^n(\tau) + \psi_{\phi\phi_2}^n(\tau) \right] e^{-s\tau} \, d\tau \tag{22}$$

$$\Psi_{oo}(s) = \frac{1}{2} e^{-b^2} \sum_{n=0}^{\infty} \frac{b^{2n}}{n!} \left[\Psi_1^{(n)}(s) + \Psi_2^{(n)}(s) \right]. \tag{23}$$

Since $\Psi_1^{(n)}(s)$ and $\Psi_2^{(n)}(s)$ have already been derived, Eq. 23 corresponds only to summing with the given weighting.

$\Psi_{oo}(s)$ can be represented by plotting the poles and residues of each of the respective terms in the s-plane.

In the example, in which the only nonzero terms are those containing only A_1 and A_{-1}, the poles and residues are found from the result above. In Fig. XIV-20 the terms of $\Psi_{oo}(s) \cdot 2e^{b^2}$ are plotted for values of n from 1 to 4. The higher terms are easily obtained by the two-dimensional technique and by multiplying the result by $b^{2n}/n!$. The term for n = 0 corresponds to an impulse at $\omega = 0$ and is omitted.

The power density spectrum of $f_o(t)$ is the Fourier transform of $\psi_{oo}(\tau)$, or $\Psi_{oo}(j\omega)$. This expression is obtained by replacing s by $j\omega$ in $\Psi_{oo}(s)$ and then multiplying by $1/2\pi$ to conform to the common definition of the Fourier transform.

$$\Psi_{oo}(j\omega) = \frac{1}{2\pi} \cdot \frac{1}{2} \cdot e^{-b^2} \sum_{n=0}^{\infty} \frac{b^{2n}}{n!} \left[\Psi_1^{(n)}(j\omega) + \Psi_2^{(n)}(j\omega) \right]. \tag{24}$$

$\Psi_{oo}(j\omega)$ can be obtained from s-plane representation. Except for a ·ultiplicative constant, $\Psi_{oo}(j\omega)$ can be obtained by taking the reciprocal of the vector f\,om the pole of a term of $\Psi_{oo}(s)$ to the point s = $j\omega$ times the residue of that pole and summing these results over all poles.

Notice that this is quite different from the normal s-plane technique of finding the product of the vectors from the zeros divided by the product of the vectors from the poles. An attempt was made to find $\Psi_{oo}(s)$ as a ratio of products of the form $(s-s_i)/(s-s_j)$. This attempt was not successful in the general case, because of the complexity arising from the infinite number of poles.

In any specific case it is not necessary to complete the sum in Eq. 24 because the factor $b^{2n}/n!$ approaches zero rapidly as n becomes large. The size of n needed depends on the value of b^2 and the degree of accuracy desired in $\Psi_{oo}(j\omega)$.

$\left(\dfrac{b^8 A_1^4}{4!}\right)$ × (upper left, $i4\omega_0$) $\left(-\dfrac{b^8 A_{-1}^4}{4!}\right)$ × (upper right, $i4\omega_0$)

$\left(\dfrac{b^6 A_1^3}{3!}\right)$ × $\left(-\dfrac{b^6 A_{-1}^3}{3!}\right)$ × ($i3\omega_0$)

$\left(4\,\dfrac{b^8 A_1^3 A_{-1}}{4!}\right)$ × $\left(\dfrac{b^4 A_1^2}{2!}\right)$ × $\left(-\dfrac{b^4 A_{-1}^2}{2!}\right)$ × $\left(-4\,\dfrac{b^8 A_{-1}^3 A_1}{4!}\right)$ × ($i2\omega_0$)

$\left(3\,\dfrac{b^6 A_1^2 A_{-1}}{3!}\right)$ × $(b^2 A_1)$ × $(-b^2 A_{-1})$ × $\left(-3\,\dfrac{b^6 A_{-1}^2 A_1}{3!}\right)$ × ($i\omega_0$)

$\left(6\,\dfrac{b^8 A_1^2 A_{-1}^2}{4!}\right)$ × $\left(2\,\dfrac{b^4 A_1 A_{-1}}{2!}\right)$ × $\left(-2\,\dfrac{b^4 A_{-1}^2 A_1}{2!}\right)$ × $\left(-6\,\dfrac{b^8 A_{-1}^2 A_1^2}{4!}\right)$ ×

Re axis: × × × | × × × →Re
$-4a \quad -3a \quad -2a \quad -a \quad 0 \quad a \quad 2a \quad 3a \quad 4a$

$\left(3\,\dfrac{b^6 A_1 A_{-1}^2}{3!}\right)$ × $(b^2 A_{-1})$ × $(-b^2 A_1)$ × $\left(-3\,\dfrac{b^6 A_{-1} A_1^2}{3!}\right)$ × ($-i\omega_0$)

$\left(4\,\dfrac{b^8 A_1 A_{-1}^3}{4!}\right)$ × $\left(\dfrac{b^4 A_{-1}^2}{2!}\right)$ × $\left(-\dfrac{b^4 A_1^2}{2!}\right)$ × $\left(-4\,\dfrac{b^8 A_{-1} A_1^3}{4!}\right)$ × ($-i2\omega_0$)

$\left(\dfrac{b^6 A_{-1}^3}{3!}\right)$ × $\left(-\dfrac{b^6 A_1^3}{3!}\right)$ × ($-i3\omega_0$)

$\left(\dfrac{b^8 A_{-1}^4}{4!}\right)$ × $\left(-\dfrac{b^8 A_1^4}{4!}\right)$ × ($-i4\omega_0$)
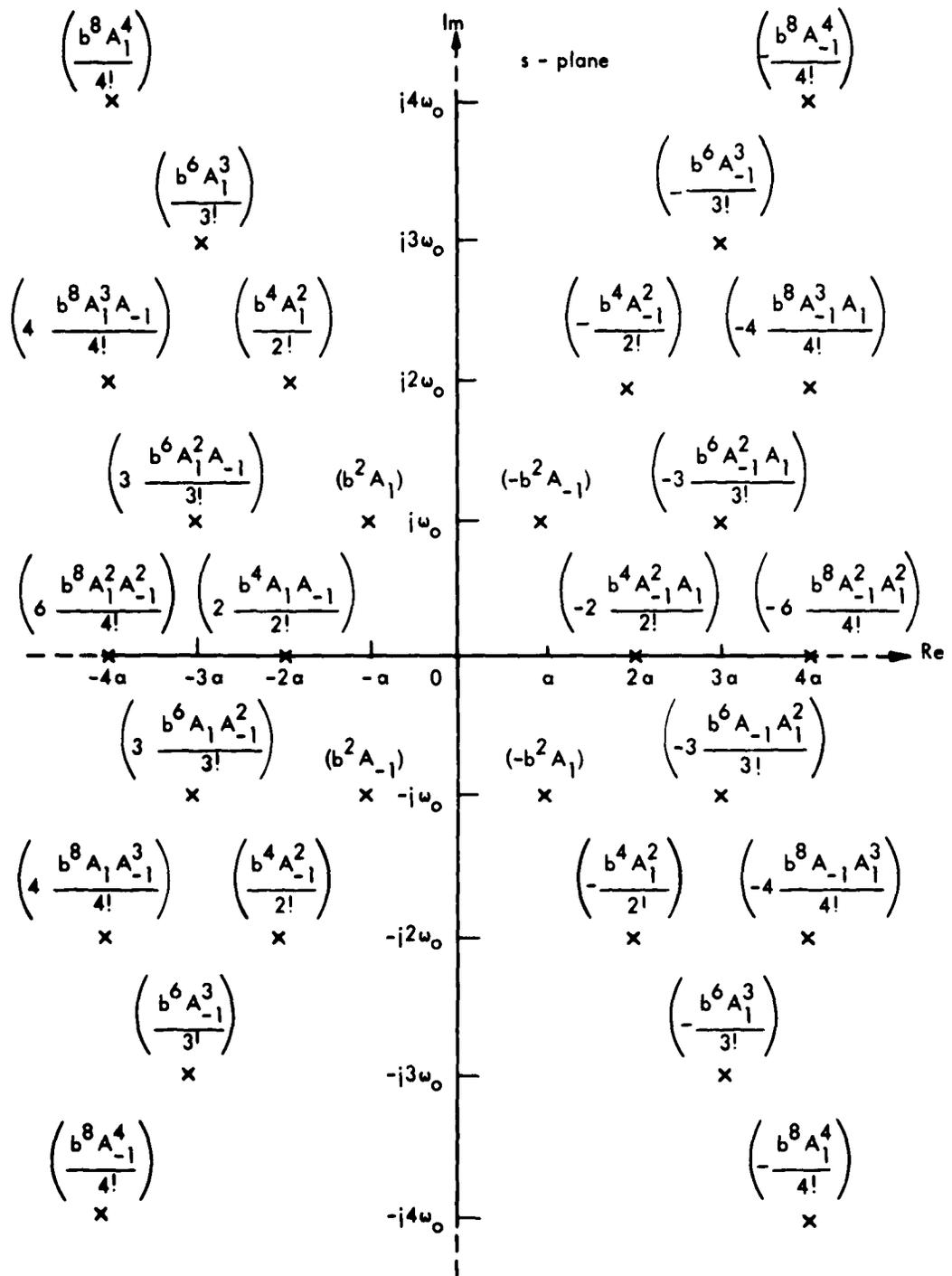
Fig. XIV-20. Terms of $2e^{b^2}\Psi_{00}(s)$.

Because of the simplicity of Eq. 24 it is a relatively simple matter to program a computer to calculate the sum for any arbitrary filter autocorrelation function. This program is now being developed at the Computation Center of the Massachusetts Institute of Technology.

3. Experiment

a. Experimental Procedure

The block diagram of the experimental system is shown in Fig. XIV-21.

The sinusoid $f_c(t)$ is phase-modulated by $s(t)$ which is the output of a shaping filter whose input is white Gaussian noise. The output of the phase modulator, $f_1(t)$, has a spectrum centered about the frequency of $f_c(t)$. In order to facilitate measurements of the spectrum it was desirable to center this spectrum about zero frequency, which is equivalent to setting the frequency of $f_c(t)$ equal to zero. This was accomplished in
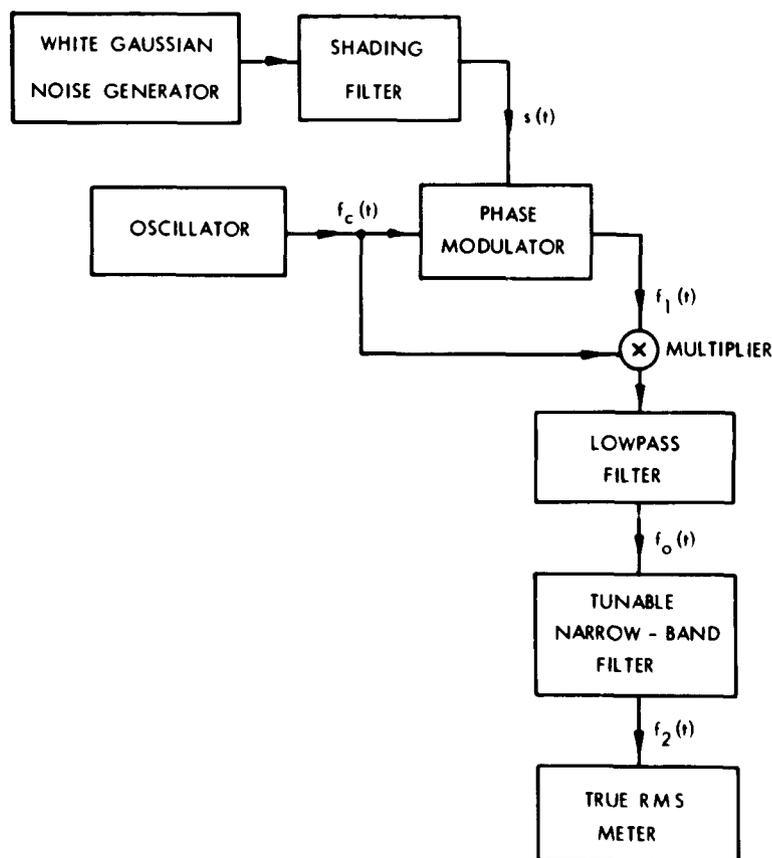
Fig. XIV-21. Experimental system.

practice by multiplying $f_1(t)$ by $f_c(t)$ and eliminating the double-frequency components with a lowpass filter. The power density spectrum of $f_0(t)$ was measured by passing $f_0(t)$ through a narrow-band filter and measuring the output on a true rms meter.

b. Implications of the Theory on Design

In order to verify the theory experimentally it was necessary to synthesize a filter that was such that when the response of this filter to white Gaussian noise was used as a signal to phase-modulate a sinusoid, the resulting power density spectrum would be easily recognizable. In order to keep the calculational difficulties to a minimum, it was desirable to use a filter with only two nonzero values of A_k in the expansion of the auto-correlation of its impulse response, as in the example used in the theoretical derivation. The resulting power density spectrum of the phase-modulated wave can then be obtained from Fig. XIV-20 by taking the reciprocal of the vector from a pole to the frequency under consideration times the residue at that pole and summing these results over all poles. From Fig. XIV-20 it can be seen that if the ratio of ω_0/a is large and if b is chosen to be approximately one, the dominant terms at $\omega = 0$ will be from the poles at $s = \pm 2a$ and the dominant terms at $\omega = \omega_0$ will be from the poles at $s = \pm a + j\omega_0$. Under these conditions, it is possible to have a power density spectrum that has a dip between $\omega = 0$ and $\omega = \omega_0$.

Since this power density spectrum has the property of being easily recognizable and has implications in the study of coupled oscillators, as shown by Wiener[1] and Schetzen,[2] a filter was designed to give this dip.

In order to make the dip in the power density spectrum prominent, it is desirable to have the magnitude at $\omega = 0$ approximately the same as the magnitude at $\omega = \omega_0$. In order to satisfy this condition, the following relation must be satisfied:

$$\frac{b^2 A_1}{a} \approx \frac{2b^4 A_1 A_{-1}}{2! \, 2a}. \tag{25}$$

It will be shown that the magnitudes of A_1 and A_{-1} are approximately 1/2. In that case

$$b \approx 2. \tag{26}$$

The symbol, b, is the product of the normalization constant of the filter and the gain of the phase modulator. The instantaneous phase of the phase-modulated signal under consideration can be expressed as Eq. 5. Since $x(t)$ is defined as white Gaussian noise with zero mean and a power density spectrum of $1/2\pi$, the mean of $g(t)$ is zero and the variance of $g(t)$ is equal to b^2.

According to Eq. 25 it is desirable to have $b \approx 2$, which implies a variance of 4 in the phase of $f_0(t)$. This then puts a difficult design constraint on the phase modulator,

since the phase must be able to swing plus or minus three times the standard deviation. The phase modulator must swing $\pm 3 \times 2 \times 57.3$ degrees or a total swing of approximately 700 degrees. If the value of the power density spectrum at $\omega = 0$ is allowed to be one-fourth the value at $\omega = \omega_o$, b could be reduced to 1, which would reduce the necessary swing in phase to 350 degrees. Most phase-modulation systems in use have a linear swing in phase of less than 90 degrees. Since a value of b greater than 1 was desired, a phase modulator had to be constructed with a swing of at least 360 degrees.

The autocorrelation function of the impulse response of the filter should have poles located at $s = \pm a \pm j\omega_o$. If the Fourier transform of a filter is $H(j\omega)$, then the Fourier transform of its autocorrelation function is $|H(j\omega)|^2$. It is easily seen, then, that the desired filter has poles at $s = -a \pm j\omega_o$. A filter with these poles is easily constructed by a parallel connection of a resistor, an inductor, and a capacitor. The input to the filter must be a current and the output a voltage.

c. Component Parts

The major part of this experiment was the design and construction of the phase modulator because of the range required. The description of the phase modulator is explained in detail in the author's thesis,[4] together with a description of the other component parts, which are shown in Fig. XIV-21.

d. Operational Data

The frequency of the sinusoid $f_c(t)$ in Fig. XIV-21 was 1500 kc. The modulating signal, g(t), was bandlimited so that there were no frequency components above 20 kc. The phase modulator could operate at well above this frequency.

Figure XIV-22 shows the experimental plot of the phase angle of the 1500-kc sinusoid versus the input-modulating voltage. The curve is very linear. It is at all points within 0.8 per cent of the best straight line drawn through the points. The curve levels off after ± 3.5 volts as a result of Zener diodes used in the construction.[4]

Because of the design of the phase modulator, the output signal is free of amplitude modulation.

The rms value of the input voltage was adjusted to be 1 volt. Since the cutoff voltage of the phase modulator is 3.5 volts, 99.9 per cent of the time the input voltage was in the linear range of the phase modulator. For this reason the probability density of the phase could be considered to be pure Gaussian.

The shaping filter had a resonant frequency of 8.3 kc and a Q of 13.5. For that reason the spectrum of the Gaussian noise generator which was level between 20 cps and 20 kc was considered white for the purposes of this experiment.
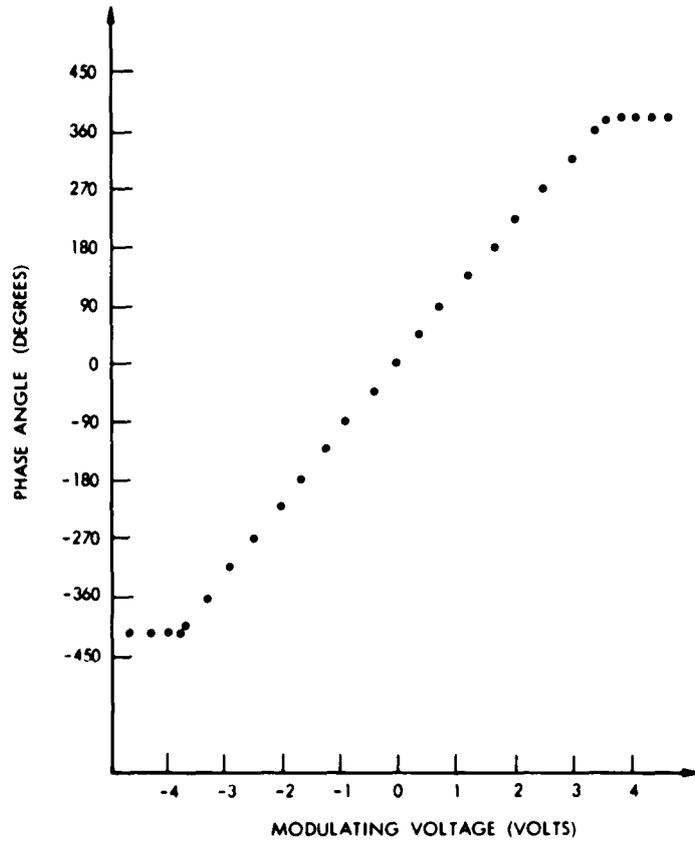
Fig. XIV-22. Phase angle of sinusoid versus input modulating voltage.

4. Comparison of Theoretical and Experimental Results

The system function of a parallel RLC circuit as described in section 3 is merely the impedance. The system function, H(s), therefore is

$$H(s) = \frac{s/C}{s^2 + \frac{s}{RC} + \frac{1}{LC}}. \tag{27}$$

By defining $a = 1/RC$, $\omega_o = 1/\sqrt{LC}$, $\omega_d = \sqrt{\omega_o^2 - a^2}$, the normalized autocorrelation function of the impulse response of the system, $\psi_{\phi\phi}(\tau)$, is

$$\psi_{\phi\phi}(\tau) = \frac{1}{2}\left(1 + j\frac{a}{\omega_d}\right) \exp[-(a-j\omega_d)|\tau|] + \frac{1}{2}\left(1 - j\frac{a}{\omega_d}\right) \exp[-(a+j\omega_d)|\tau|]. \tag{28}$$

In the notation of the example discussed in section 2, $A_1 = \frac{1}{2}\left(1 + j\frac{a}{\omega_d}\right)$, $A_{-1} = \frac{1}{2}\left(1 - j\frac{a}{\omega_d}\right)$,

$s_1 = -a + j\omega_d$, and $s_{-1} = -a - j\omega_d$.

As stated in section 2, the power density spectrum of the phase-modulated wave, $f_o(t)$, is obtained by means of Fig. XIV-20. This is obtained by taking the reciprocal of the vector from a pole to the frequency under consideration times the residue at that pole and summing over all poles. Consider one of the poles in Fig. XIV-20 and its mirror reflection across the imaginary axis. If a vector is drawn from each of these poles to a point on the imaginary axis, both vectors will have the same imaginary component and the real components will be negatives of each other. Because of this fact and since the residues of these poles are the negative complex conjugates of each other, the sum of the contributions from this pair of poles will be twice the real part of the contribution from either one of the poles. For this reason it is possible to consider only one-half of the poles, say the poles in the left-half plane, by taking twice the real part of each contribution.

Figure XIV-23 shows the notation used to designate the contributions from each of the specific poles. The sum contribution from the poles at $s = \pm na + jm\omega_d$ is labeled CUnm and the contribution from the poles at $s = \pm na - jm\omega_d$ is labeled CLnm. CU11 is therefore the sum contribution from the poles at $s = \pm a + j\omega_d$ or twice the real part of the contribution from the pole at $s = -a + j\omega_d$.

$$CU11 = \frac{1}{2\pi} \cdot \frac{1}{2} e^{-b^2} 2\text{Re} \frac{b^2 A_1}{(j\omega + a - j\omega_d)} = \frac{1}{2\pi} \cdot \frac{1}{2} e^{-b^2} \frac{b^2 \left[a + \frac{a}{\omega_d}(\omega - \omega_d)\right]}{a^2 + (\omega - \omega_d)^2}. \tag{29}$$

The contribution from the poles at $s = \pm a - j\omega_d$ is

$$CL11 = \frac{1}{2\pi} \cdot \frac{1}{2} e^{-b^2} \frac{b^2 \left[a - \frac{a}{\omega_d}(\omega + \omega_d)\right]}{a^2 + (\omega + \omega_d)^2}. \tag{30}$$

The contribution from the poles at $s = \pm 2a + j2\omega_d$ is

$$CU22 = \frac{1}{2\pi} \cdot \frac{1}{2} e^{-b^2} 2\text{Re} \frac{\frac{b^4}{2!} A_1^2}{j\omega + 2a - j2\omega_d} = \frac{1}{2\pi} \cdot \frac{1}{2} e^{-b^2} \frac{b^4}{4} \frac{\left[2a\left(1 - \frac{a^2}{\omega_d^2}\right) + \frac{2a}{\omega_d}(\omega - \omega_d)\right]}{4a^2 + (\omega - \omega_d)^2}. \tag{31}$$

The contributions from other poles are derived in the same manner and will not be listed here.

The number of poles to be considered in any one case depends on the accuracy desired in $\Psi_{oo}(j\omega)$ and in the values of b, a, and ω_d.
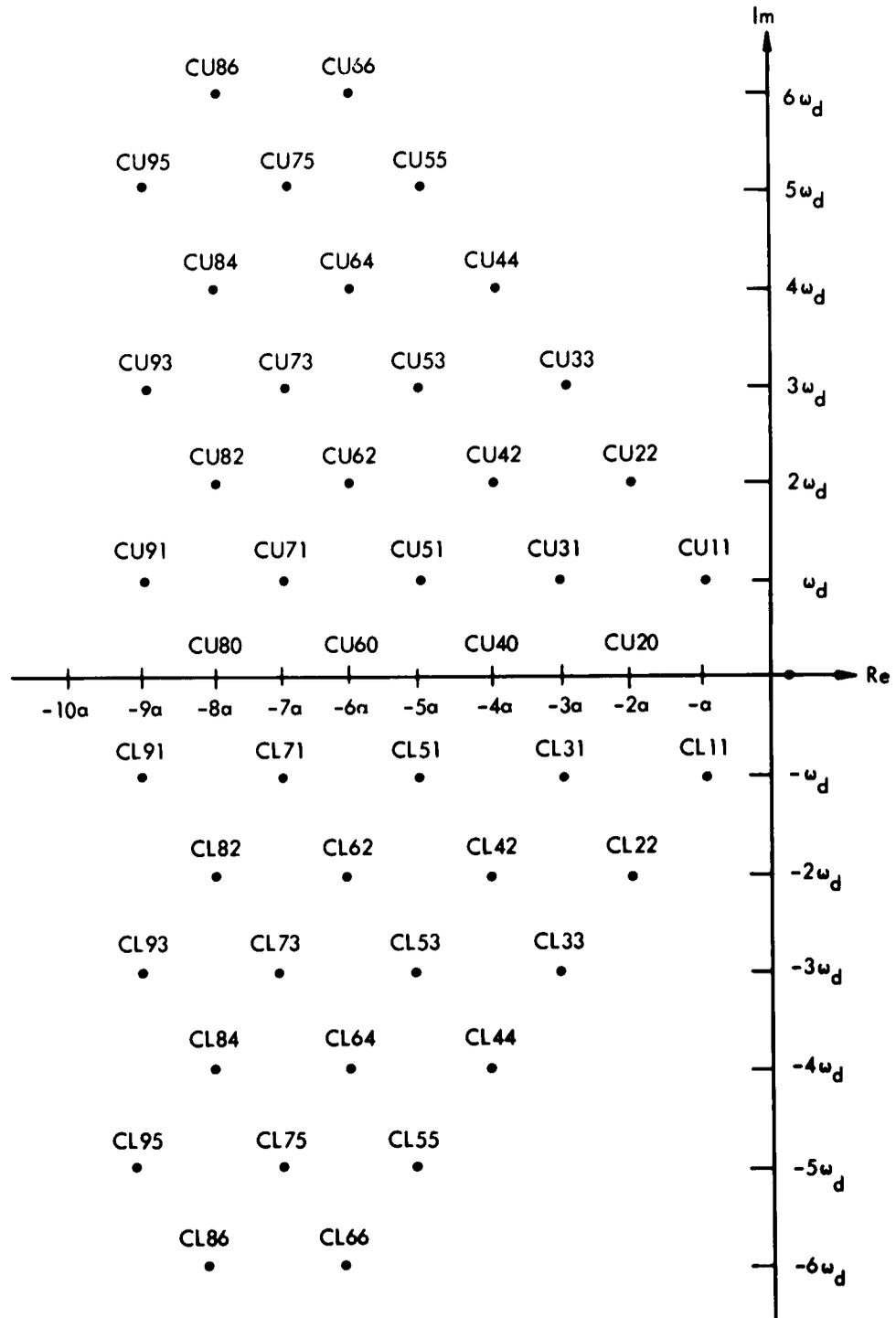
Fig. XIV-23. Notation used to designate contributions from each of the specific poles.

The value of b is the standard deviation of phase of $f_1(t)$ in radians. The measured rms value of the signal, s(t), applied to the phase modulator in this experiment was 1.05 volts. The standard deviation of the phase can be obtained from the phase versus voltage curve, Fig. XIV-22. The slope of this curve is 105 degrees per volt. Therefore the standard deviation is $(1.05) \cdot (105) = 110.3$ degrees or 1.926 radians, and the value of b used in the calculations is 1.926. The measured Q of the filter was 13.5, which yields a value of 27 for the ratio ω_d/a. The measured resonance of the filter was 8.3 kc; therefore $\omega_d \approx 2\pi \cdot 8.3 \times 10^3$ radians per second and $a = 1.930 \times 10^3$ nepers per second.

With these values of b, a, and ω_d used, the contributions from the various poles were calculated for some specific frequencies. The variable used in these calculations is f, where $\omega = 2\pi f$. These results are shown in Table XIV-1. Only the contributions from the largest poles are shown. The contributions from the other poles were too small to be of interest in the range of frequencies under consideration. The sum of these contributions, labeled CT in Table XIV-1, is plotted in Fig. XIV-24 as a solid line. It is important to notice that the values of b, a, and ω_d chosen do, indeed, produce the desired dip in the spectrum between $\omega = 0$ and $\omega = \omega_d$.

The experimental measurements were made by applying the phase-modulated signal, $f_0(t)$, to a narrow-band filter. The filter was tuned to the desired frequency and the output of the filter was measured on a true rms meter. The square of this measurement was recorded.

Even though the meter had an effective time constant of approximately 5 sec, the meter reading fluctuated back and forth a distance that was a sizable percentage of the average. In order to overcome this result of finite time integration, the meter reading was averaged by eye over several minutes. The results of these measurements are shown as bars superimposed over the theoretical curve in Fig. XIV-24. The length of the bar indicates the range in which the true average could lie on the basis of the measurements and the center is the estimated average value. Since the amplitude of the modulated sinusoid was not one, and since the gain of the narrow-band filter was not one, the absolute magnitude of the series of measurements was not known. For that reason the experimental curve was multiplied by a constant that seemed to yield the best overall fit to the theoretical curve.

Figure XIV-24 shows that the experimental data follow the theoretical curve very closely. It can be seen that both the experimental and the theoretical curves have the desired dip between $\omega = 0$ and $\omega = \omega_d$. It can be noted further that the dip occurs at the same place in both curves and that relative heights of the peaks also agree. In fact, there are only a few minor departures of the experimental curve from the theoretical curve.

Some possible explanations for these minor departures might be errors arising from

Table XIV-1. Contributions of the poles in Fig. XIV-23 at various frequencies in units of 10^{-9} watts per rad/sec.

F	CU11	CL11	CU22	CU20	CL22	CU31	CL31	CU42	CU40	CU53	CL51	CU62	CU60	CU71	CU80	GT
0	0	0	-2	3476	-2	17	17	0	1497	20	20	1	382	8	62	5495
500	0	0	-2	2092	-2	20	15	0	1284	22	17	1	356	9	59	3873
1000	1	0	-2	953	-2	24	13	0	901	26	15	1	295	11	53	2288
1500	1	-1	-2	500	-2	28	11	1	602	30	14	2	230	12	45	1470
2500	3	-1	-2	198	-2	44	9	1	292	42	11	2	135	17	30	774
3500	6	-1	-2	104	-2	62	7	2	164	61	9	3	83	23	20	539
4500	13	-1	-2	64	-2	101	5	3	104	95	7	3	55	34	14	493
5500	30	-1	-2	43	-2	187	4	4	72	160	6	4	39	54	10	606
6500	83	-1	-2	31	-2	444	3	6	52	299	5	5	28	87	8	1016
7000	167	-1	-1	27	-2	682	3	7	45	418	5	6	25	109	7	1495
7500	435	-1	-1	23	-2	1187	3	8	39	571	4	7	22	132	6	2433
8000	1848	-1	-1	20	-2	1920	2	10	35	708	4	8	19	148	5	4723
8300	3743	-1	0	19	-2	2449	2	11	32	740	4	9	18	152	5	6880
9000	656	-1	0	16	-2	1402	2	14	27	623	3	11	15	139	4	2911
9500	284	-1	1	14	-2	836	2	17	25	473	3	13	14	118	4	1761
10500	91	-1	2	12	-2	350	1	25	20	256	3	18	11	77	3	868
11500	47	-1	5	10	-2	186	1	38	17	149	2	26	10	50	3	546
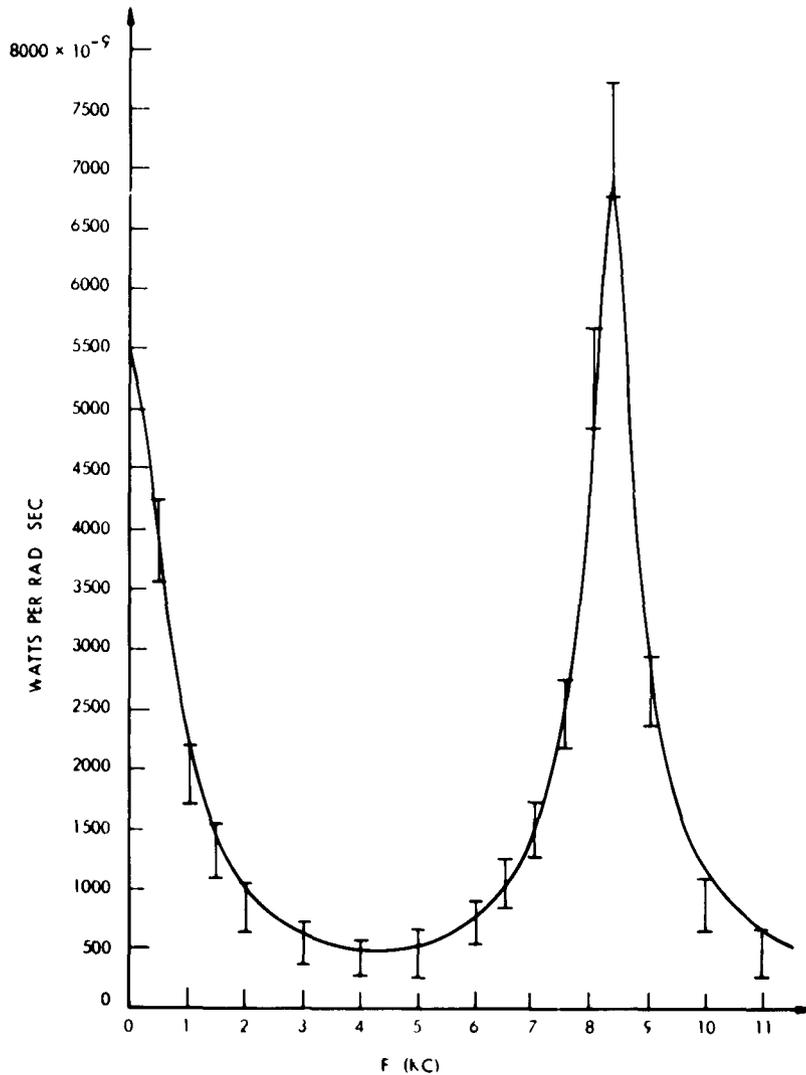
Fig. XIV-24. Theoretical and experimental curves.

the finite time averaging and possibly a value for b, the rms phase deviation, that is slightly in error and affects the spectrum greatly. Another possible explanation of the departures might be that the noise source is not a true Gaussian source. Even though lower-order moments are those of a true Gaussian source, departures in the higher-order moments would cause experimental errors.

It would be interesting to measure the power-density spectrum by using other values of b. As observed, a higher value of b would make stronger use of the higher moments of the noise waveform. Departures of the higher-order moments from those of a true

Gaussian wave would increase the discrepancy between the measured spectrum and the theoretical calculation because the latter is based on a true Gaussian source. These measurements could be used for checking the higher moments of a Gaussian source.

In many current studies, the power density spectrum of $e^{j\theta(t)}$, where $\theta(t)$ is a non-linear operation on a Gaussian process, is of interest. The theoretical calculations in these studies become very complicated (see, for example, the calculations of Schetzen[2] in which the operation is the sum of a linear and a quadratic operation). The power density spectrum in these complicated cases could be obtained very easily by using the phase modulator described in the author's thesis.[4] The linear filter could be replaced by the desired nonlinear filter and the power density spectrum could be measured as described here.

J. K. Clemens

References

1. N. Wiener, Nonlinear Problems in Random Theory (The Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1958).

2. M. Schetzen, Some Problems in Nonlinear Theory, Technical Report 390, Research Laboratory of Electronics, M.I.T., July 6, 1962.

3. M. Schetzen, Unpublished Notes, 1960.

4. J. K. Clemens, A Study of Phase-Modulation with a Gaussian Signal, S. M. Thesis, Department of Electrical Engineering, M.I.T., May 13, 1963.

5. Y. W. Lee, Statistical Theory of Communication (John Wiley and Sons, Inc., New York, 1960).

XV. PROCESSING AND TRANSMISSION OF INFORMATION[*]

Prof. R. M. Fano
Prof. R. G. Gallager
Prof. F. C. Hennie III
Prof. E. M. Hofstetter
Prof. D. A. Huffman
Prof. I. M. Jacobs
Prof. A. M. Manders
Prof. W. F. Schreiber
Prof. C. E. Shannon
Prof. J. M. Wozencraft
Dr. C. L. Liu
T. G. Adcock
T. M. Anderson

P. G. Arnold
M. H. Bender
E. R. Berlekamp
D. G. Botha
J. E. Cunningham
H. Dym
P. M. Ebert
D. Ecklein
D. D. Falconer
E. F. Ferretti
G. D. Forney, Jr.
R. L. Greenspan
U. F. Gronemann
A. R. Hassan

J. L. Holsinger
T. S. Huang
N. Imai
R. J. Purdy
J. F. Queenan
J. E. Savage
J. R. Sklar
K. D. Snow
I. G. Stiglitz
W. R. Sutherland
O. J. Tretiak
W. J. Wilson
H. L. Yudkin

A. PICTURE PROCESSING

1. A MEDIUM-SPEED MEDIUM-ACCURACY MEDIUM-COST ANALOG-TO-DIGITAL CONVERTER

An analog-to-digital converter has been constructed. Some of the performance specifications are listed below. Anyone who is interested may obtain a copy of the circuit diagrams.

a. General Specifications

Input: 0 to -10 volts

Output: 8 bits, natural binary code, parallel output

Conversion time: 45 μsec

b. General Description

A conversion is initiated by an external pulse. The intervals between these pulses must be greater than 45 μsec. The conversion is done on a bit-by-bit basis.

c. Construction Details

Most of the circuitry is built with medium-speed switching transistors and general-purpose diodes. The internal voltages are derived from regulated +34-volt and -25-volt power supplies.

The circuit takes up five plug-in cards. The estimated time for wiring is approximately one man-week.

An approximate list of components includes: 52 pnp switching transistors, 46 general-purpose diodes, 24 special diodes and transistors, 150 resistors (1/2 watt 5 per cent), 18 precision resistors, 20 power resistors, 5 miniature variable resistors, 61 miscellaneous capacitors, 5 electrolytic capacitors, and 7 Zener diodes.

If the components are ordered in reasonable quantities, the total cost would be less than $200.

O. J. Tretiak

2. PICTURE RECORDING AND REPRODUCING EQUIPMENT

The scanner for recording and playing back pictures has been improved on by the addition of a feedback circuit to control the light output of the cathode-ray tube. The light intensity is sensed by a multiplier phototube and compared with the input signal. The error signal is integrated over the duration of a signal pulse, amplified, and fed to the cathode-ray tube.

This arrangement serves two purposes: (a) keeping the light output uniform (for constant signal) over the entire surface of the tube, that is, to eliminate phosphor noise; and (b) during playback, making the light output a linear function of the input signal (which is important for color reproduction by linear addition of primary colors).

The effectiveness of the circuit for purpose (a) is demonstrated by the measurement of a light-intensity variation in the ratio 4:1 for a simulated phosphor response variation in the ratio 45:1.

The linearity of the output is true within 10 per cent in a large-to-small signal amplitude ratio of approximately 100:1 (which is the useful dynamic range of a typical transparency) and should be compared with the highly exponential characteristic of the cathode-ray tube alone.

U. F. Gronemann

3. IMAGE CORRECTION-TRANSMISSION EXPERIMENTS

Usually only small areas of a motion picture change significantly from frame to frame. An obvious exception occurs if the camera is moving relative to a fixed background. Since this is true, a possible way of permitting the reconstruction of a sequence of moving picture frames is the transmission of information concerning the parts of the picture which change significantly.

To determine how such a system would perform, a program was written for the TX-2 computer (located at Lincoln Laboratory, M. I. T.) to simulate the transmission process. The input data were scanned from 35-mm film and recorded on digital computer tape. On this tape pictures consisting of 128 X 128 sample points, each having one of 256 possible brightness levels, were recorded.

(a)

(b)

(c)

(d)

(e)

(f)

Fig. XV-1. Picture build-up examples.

Fig. XV-2. Still pictures of scene transitions.

The simulations were performed under the assumption of error-free transmission and sufficient buffer storage at the transmitter and receiver to allow a signal match, as well as storage of complete pictures as required. A description of the operation of the transmission system follows.

The transmitter retains a copy of the picture that the receiver has in its memory; when confronted with the new picture data that are to be transmitted, the transmitter performs calculations to determine the N points that have the N largest brightness difference in the new pictures and the copy of the receiver picture. The transmitter then proceeds to correct the brightness values of these N points; also, each time a point is corrected, the correction value is averaged into its surround by means of a decaying interpolation function. Since the interpolation process usually changes the value of the brightness differences, these differences are recalculated each time that a correction is sent to prevent sending unnecessary corrections. The process of correcting points is carried out in an interlaced pattern that requires 4 passes to check every point in the picture. If the process completes the 4 passes without having sent all N corrections, it starts over with a lower difference threshold so that the full quota of corrections may be sent.

Since there is a constant number of corrections sent for each frame, there is no queing problem when a change of scene occurs. In Fig. XV-1 is shown an example of a picture build-up from a blank screen with the value N = 1024 used, which is equal to 1/16 of the points in the picture. The first four pictures are the frames representing the 4 frames after a blank screen; the fifth and sixth pictures are representative of the quality achieved after approximately 32 frame times. Figure XV-2 shows an example of a scene transition with the same number of corrections per frame. The six frames represent 3/8 second when viewed at the 16-frame/second rate. Since information concerning the location of these points, as well as the brightness values, must be sent, the reduction ratios are not 16:1 but approximately 6:1. This calculation is based on the upper limit of information required if all corrections are sent independently.

We plan to extend this work, and to introduce certain topological constraints on the interpolation process.

<div align="right">J. E. Cunningham</div>

B. DISTANCE PROPERTIES OF TREE CODES

In this report we show that for any two positive integers r and s such that r divides s there exists a "good" tree code in the Galois field of order p^s with p^r branches per node. The code is good in the sense that an appropriately defined minimum distance criterion may be achieved which is numerically identical to that presented by Peterson[1] for the general parity-check code. This result is presented in the form of a theorem.

In the course of the proof, several lemmas are established which in themselves describe properties of tree codes which are thought to be of interest. Finally, it is shown that a "good" tree code may be transformed into a canonic form without destroying its distance properties. We shall first develop suitable notation.

Consider the n-tuple $(g_1, \ldots, g_n)$ whose coordinate entries are chosen from a Galois field of order p^s (G. F.$[p^s]$), where p is a prime and s is a positive integer. Suppose, also, that $g_1 \neq 0$ and that $n = kn_0$, where k and n_0 are positive integers. We are interested in the set of n-tuples generated by certain linear combinations of the following basic set of n-tuples.

g_1	g_2	$\cdots$	g_{n_0}	g_{n_0+1}	g_{n_0+2}	$\cdots$		g_{2n_0+1}	$\cdots$	$g_{(k-1)n_0+1}$	$\cdots$	g_{kn_0}
0	0	$\ldots$	0	g_1	g_2	$\cdots$		$\cdots$	$\cdots$	$\cdots$	$\cdots$	$g_{(k-1)n_0}$
0	0	$\ldots$	0	0	0	$\ldots$	0	g_1	$\cdots$	$\cdots$	$\cdots$	$g_{(k-2)n_0}$
$\vdots$												$\vdots$
0	0	$\ldots$	0	0	0	$\ldots$	0	0	$\ldots$	0	g_1	$\cdots$	g_{n_0}

It is sometimes convenient to represent n-tuples in functional form. Thus we shall represent the n-tuple $(g_1, \ldots, g_n)$ by the function $g(t)$ defined as

$$g(t) = \begin{cases} g_i & \text{if } t = i \quad \text{for } i = 1, 2, \ldots, n \\ 0 & \text{all other } t \end{cases}$$

Correspondingly, the array of n-tuples presented above may, at least for $t = 1, \ldots, n$, be represented by the set of functions

$$g(t), \quad g(t-n_0), \quad \ldots, \quad g(t-(k-1)n_0).$$

Consider the set of functions (n-tuples) of the form[2]

$$p(t) = x_0 g(t) + x_1 g(t-n_0) + \ldots + x_{k-1} g(t-(k-1)n_0),$$

where the coefficients of the functions $g(t-in_0)$ for $i = 0, \ldots, k-1$ are chosen from a subset E of the Galois field. If E contains m elements, there are m^k such distinct functions, for the condition $g(1) \neq 0$ implies that the set of functions $g(t-in_0)$ $i = 0, \ldots, k-1$ is linearly independent.[3]

We shall refer to the set of functions generated this way as the tree code or convolutional code generated by $g(t)$ and E. The name "tree code" derives from the fact that the elements (functions, code words, paths) of the code can be represented diagrammatically in the form of a tree. This is illustrated in Fig. XV-3 for a generator of length 6

(a)

(b)
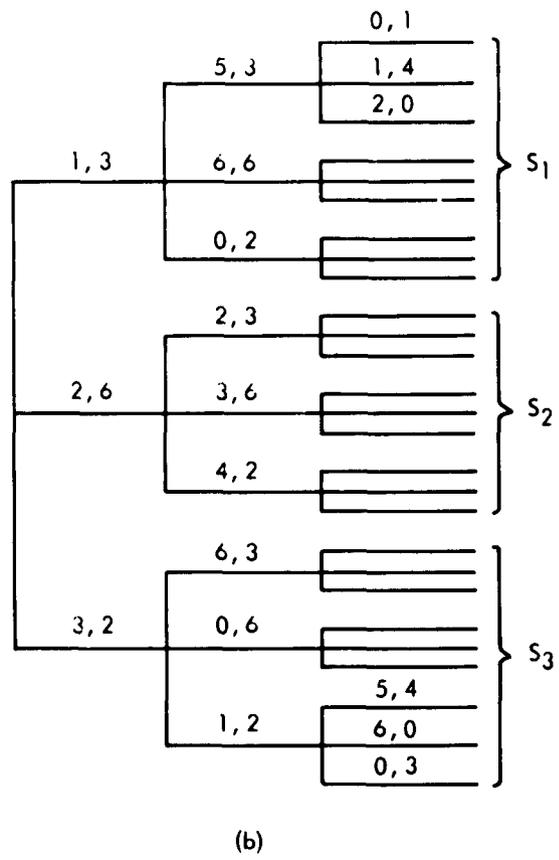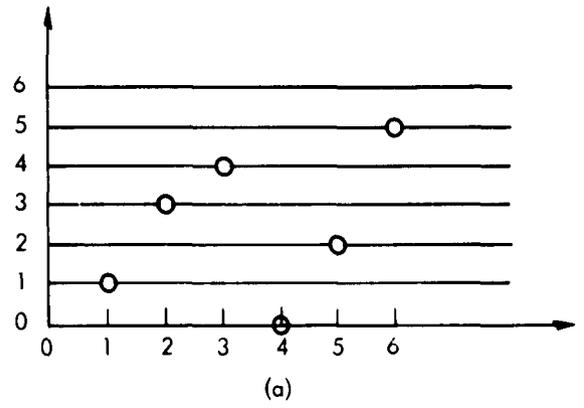
Fig. XV-3.   (a) Generator function g(t).
(b) Tree code generated by g(t) and
the set of field elements 1, 2, 3.

with entries drawn from the Galois field of order 7. In this example it is assumed also that $n_o = 2$, and that the set F consists of the field elements 1, 2, and 3.

Returning to the general case, we denote the m elements of E as e_o, e_1, ..., e_{m-1}, and consider the subsets of functions (paths):

$$S_o = \left\{ p(t) \middle| p(t) = e_o g(t) + \sum_{i=1}^{k-1} x_i g(t-in_o) \right\}$$

$$\vdots$$

$$S_{m-1} = \left\{ p(t) \middle| p(t) = e_{m-1} g(t) + \sum_{i=1}^{k-1} x_i g(t-in_o) \right\},$$

where the x_i vary over all possible values of E.

In effect, we have partitioned the totality of paths into subsets with the same initial prefix. (See, for example, the path sets S_1, S_2, and S_3 in Fig. XV-3b.) Clearly, each subset S_i contains m^{k-1} paths.

We define the distance between any pair of distinct path sets S_i, S_j as

$$D(S_i, S_j) = \min_{\substack{p(t) \in S_i \\ h(t) \in S_j}} d(p(t), h(t)),$$

where $d(p(t), h(t))$ is the conventional Hamming metric,

$$d(p(t), h(t)) = \sum_{i=1}^{n} \| p(i) - h(i) \|.$$

Here, for any element x in the Galois field

$$\| x \| = 1 \qquad \text{if } x \neq 0$$

$$\| x \| = 0 \qquad \text{if } x = 0.$$

We define the weight of a vector to be equal to the distance of the vector from the all-zero vector.

If we let

$$\mathscr{D} = \min_{\substack{i, j \\ i \neq j}} D(S_i, S_j),$$

we can now state the theorem, which is the main result of this report.

THEOREM: If E is a subfield of the Galois field of order q, then there exists a generated tree code for which $\mathscr{D}$ exceeds Δ, where Δ is the largest

integer satisfying the inequality

$$\sum_{j=0}^{\Delta-2} \binom{n-1}{j}(q-1)^j < \frac{q^{n-1}}{m^{k-1}}.$$

REMARK 1: For q = m, this condition is identical to the one presented by Peterson for the general parity-check code.[1]

PROOF: We shall prove the theorem in a series of lemmas that serve as convenient guideposts to the main thread of the argument. After the proof, we shall discuss the effect of weakening the hypothesis by requiring only that E be an additive subgroup of the Galois field and not necessarily a subfield. This is of interest, since the requirement that E be a subfield is somewhat restrictive. The only subfields of a Galois field of order p^m are of order p^r, where r divides m.[4]

LEMMA 1: $D(S_i, S_j)$, $i \neq j$, is independent of the choice of i and j.

By definition,

$$D(S_i, S_j) = \min_{\substack{p(t) \in S_i \\ h(t) \in S_j}} d(p(t), h(t)).$$

But functions belonging to S_i and S_j must have respectively the forms

$$e_i g(t) + \sum_{l=1}^{k-1} x_l g(t-l n_o)$$

and

$$e_j g(t) + \sum_{l=1}^{k-1} y_l g(t-l n_o),$$

in which each of the 2(k-1) coefficients x_l and y_l is chosen from E. Thus

$$D(S_i, S_j) = \min_{\substack{\{x_l\} \\ \{y_l\}}} \left\{ \sum_{v=1}^{n} \left\| (e_i - e_j) g(v) + \sum_{l=1}^{k-1} (x_l - y_l) g(v-l n_o) \right\| \right\}.$$

Clearly, however, since x_l and y_l are chosen from E, which is a subfield and therefore an additive group, we have

$$D(S_i, S_j) = \min_{\{x_l\}} \left\{ \sum_{v=1}^{n} \left\| (e_i - e_j) g(v) + \sum_{l=1}^{k-1} x_l g(v-l n_o) \right\| \right\}.$$

Furthermore, $(e_i - e_j) \neq 0$ implies the existence of a multiplicative inverse $(e_i - e_j)^{-1}$, and thus

$$D(S_i, S_j) = \min_{\{x_l\}} \left\{ \sum_{v=1}^{n} \left\| g(v) + \sum_{l=1}^{k-1} (e_i - e_j)^{-1} x_l g(v - l n_0) \right\| \right\}.$$

But E, a field, implies that

$$\left\{ (e_i - e_j)^{-1} x \right\}_{x \in E} = \{x\}_{x \in E}.$$

Hence

$$D(S_i, S_j) = \min_{\{x_l\}} \left\{ \sum_{v=1}^{n} \left\| g(v) + \sum_{l=1}^{k-1} x_l g(v - l n_0) \right\| \right\}.$$

The right-hand side is independent of the choice of i and j, $i \neq j$, and thus Lemma 1 is proved.

Since E is a field, it must contain a zero element. It is notationally convenient to assume that, in fact, $e_0 = 0$ and, correspondingly, S_0 is the path set containing the all-zero n-tuple. We shall refer to S_0 as the zero-path set. Adopting this notation, we note that the reasoning used to establish Lemma 1 proves almost directly the following lemma.

LEMMA 2: The quantity $\mathscr{D}$ is equal numerically to the weight of the minimum-weight path in the code which does not belong to S_0.

LEMMA 3: The number of distinct n-tuples that, in conjunction with the subfield E, generate the same code is exactly equal to $(m-1)m^{k-1}$.

Assume a code generated by the function g(t). By definition, $g(1) \neq 0$. The number of paths in this code which satisfy the constraint $p(1) \neq 0$ is equal to $(m-1)m^{k-1}$. Any such path, however, may be used to generate the same code. That is,

$$\left\{ \sum_{i=0}^{k-1} x_i p(t - i n_0) \right\}_{x_i \in E} = \left\{ \sum_{i=0}^{k-1} x_i g(t - i n_0) \right\}_{x_i \in E}.$$

In words, each and every one of the k^m n-tuples that may be expressed as a linear combination of g(t) and its translates may also be expressed as a linear combination of p(t) and its translates. We have thus established the fact that the number of distinct n-tuples that generate the same code is greater than or equal to $(m-1)m^{k-1}$.

However, any code generator must itself belong to the code, and, furthermore, cannot belong to S_0. This establishes the reverse inequality, and hence Lemma 3 is proved.

We shall say for the purposes of this report that two tree codes are essentially distinct if the only paths that they have in common belong to their respective zero path sets. We now formulate

LEMMA 4: Two generated tree codes are either essentially distinct or identical.

Assume two generated tree codes C_1 and C_2 that are not essentially distinct. Then there exists a path $p(t)$ with nonzero first coordinate belonging to both C_1 and C_2. But such a path can serve as a generator both for C_1 and C_2. It follows, therefore, that C_1 must be identical to C_2.

LEMMA 5: The number of essentially distinct tree codes that can be generated with generators of length n in a Galois field of order q is exactly equal to $\dfrac{(q-1)q^{n-1}}{(m-1)m^{k-1}}$ where m is the order of the subfield E.

Initially, we note that if $q = p^s$, then E is a subfield only if $m = p^r$, where r divides s. This ensures that, in fact, $\dfrac{(q-1)q^{n-1}}{(m-1)m^{k-1}}$ is a positive integer.

The number of distinct code vectors in G. F.$[q]$ with a nonzero first coordinate is equal to $(q-1)q^{n-1}$. Consider the code, say C_1, which is generated by one such vector. In C_1 there will be a total of $(m-1)m^{k-1}$ vectors with nonzero first coordinates. If $(q-1)q^{n-1} - (m-1)m^{k-1} > 0$, there exists an n-tuple with a nonzero first coordinate that is not in C_1. Consequently, we can generate a second code C_2 that is, by Lemma 4, essentially distinct from C_1. By proceeding in this fashion, it is clear that the number of essentially distinct codes that we can generate is exactly equal to $\dfrac{(q-1)q^{n-1}}{(m-1)m^{k-1}}$.

The final argument in the proof of the theorem is basically a comparison between the number of low-weight n-tuples with a nonzero first entry and the number of essentially distinct codes. Lemma 2 implies that $\mathscr{D}$, the minimum distance between any pair of distinct path sets in a code, is equal to the weight of the minimum-weight path in the code which does not belong to the zero subset. The total number of n-tuples of weight less than or equal to $\Delta-1$, which have a nonzero first coordinate, is

$$(q-1) \sum_{j=0}^{\Delta-2} \binom{n-1}{j}(q-1)^j.$$

In each code there will be at least $m-1$ code words of the same weight. Thus the total number of essentially distinct codes that can contain a path (in a nonzero subset) of weight less than Δ is, at most,

$$\left(\frac{q-1}{m-1}\right) \sum_{j=0}^{\Delta-2} \binom{n-1}{j}(q-1)^j.$$

Consequently, as long as the total number of essentially distinct codes exceeds this number, there will always exist a code with $\mathscr{D} \geq \Delta$. The theorem now follows from Lemma 5.

The proof of the theorem depends essentially on Lemma 4, which guarantees that two codes cannot partially overlap in a nontrivial way. More precisely, assume a code C_1 and an n-tuple $p(t)$ with $p(1) \neq 0$ with the property that $p(t)$ does not belong to C_1. Then Lemma 4 implies that the code generated by $p(t)$ is essentially distinct from C_1. This is no longer true necessarily if E is only restricted to be an additive subgroup of the Galois field.

Consider, for example,[5] the Galois field of order 16 which can be represented as the field of polynomials over G. F.[2] with multiplication modulo the polynomial $x^4 + x + 1$. The four elements 1, x, 1 + x, 0 form an additive subgroup of G. F.[16]. Assume a generator whose first entry is the Galois-field element $1 + x + x^3$. The corresponding first entries in the four branches stemming from the first node will be

$$1(1+x+x^3) = 1 + x + x^3$$

$$x(1+x+x^3) = 1 + x^2$$

$$(1+x)(1+x+x^3) = x + x^2 + x^3$$

$$0(1+x+x^3) = 0.$$

Now consider a second generator whose first entry is the element $x^2 + x^3$. Such a generator clearly cannot belong to the code generated by the first generator. Yet

$$x(x^2+x^3) = 1 + x + x^3,$$

which coincides with one of the first-branch entries of the first code. It follows that by appropriate juggling it is possible to construct a pair of generators $g_1(t)$ and $g_2(t)$ so that $g_2(t)$ does not belong to the code generated by $g_1(t)$, which we designate C_1. But the code generated by $g_2(t)$, C_2, is still not essentially distinct from C_1. That is, there exists a path $p(t)$ with $p(1) \neq 0$ which is common to both codes. Thus Lemma 4 is no longer true if the assumption that E is a subfield is replaced by the weaker assumption that E is an additive subgroup. Whether or not the theorem is true under these weaker conditions is not yet resolved — at least to the author's knowledge. In this direction we might point out that Lemma 2 can be shown to be true, although Lemma 1 is false, under the assumption that E is an additive subgroup.

We have shown that for any two positive integers r and s such that r divides s there exists a generator for a tree code in the Galois field of order p^s with p^r branches per node, which is good in the sense that an appropriately defined minimum distance criterion can be satisfied. Furthermore, the proof is constructive in the sense that by

simply choosing generators and testing the resulting codes one is guaranteed that ulti-mately a good code will be found.[6] Let us suppose that such a code has, in fact, been found with generator function $g(t)$.

REMARK 2: (a) The code can be implemented by using only modulo p addition and multiplication.

(b) From such a code one can derive a code having the same distance properties which is in canonic form (to be described below).

We proceed to discuss Remark 2a.

A field of order p^r is a vector space of dimension r over the prime field of order p. We thus can select r elements from F, say $e_1, \ldots, e_r$ such that every element in E may be written in the form

$$\lambda_1 e_1 + \ldots + \lambda_r e_r$$

for some choice of $\lambda_1, \ldots, \lambda_r$ where $\lambda_1, \ldots, \lambda_r$ are elements in G. F.$[p]$.

Consider the multiples of the generator function $e_1 g(t), \ldots, e_r g(t)$ (where multipli-cation is performed according to the rules of G. F.$[p^s]$), and designate the corresponding products $g_1(t), \ldots, g_r(t)$.

With each such function $g_1(t), \ldots, g_r(t)$ we can associate an n-tuple

$$
\begin{matrix}
a_{11} & \cdots & a_{1n} \\
\vdots & & \vdots \\
a_{r1} & \cdots & a_{rn}
\end{matrix}
$$

where the a_{ij} are elements of G. F.$[p^s]$.

The Galois field of order p^s is, however, a vector space of dimension s over the prime field of order p. Correspondingly, we, in fact, can associate with each function $g_1(t), \ldots, g_r(t)$ an sn-tuple of prime-field elements

$$
\begin{matrix}
b_{11} & \cdots & b_{1(sn)} \\
\vdots & & \vdots \\
b_{r1} & \cdots & b_{r(sn)}
\end{matrix}
$$

where the b_{ij} are elements of G. F.$[p]$.

Correspondingly, letting $\hat{g}_i(t)$ denote the expansion of $g_i(t)$ in prime-field elements, we can describe our code as the set of sn-tuples of the form

$$\sum_{i=0}^{k-1} \lambda_{1i}\hat{g}_1(t-isn_o) + \sum_{i=0}^{k-1} \lambda_{2i}\hat{g}_2(t-isn_o) + \ldots + \sum_{i=0}^{k-1} \lambda_{ri}\hat{g}_r(t-isn_o),$$

where the λ_{ij} are elements of the prime field.

In this formulation all addition and multiplication is done modulo p. Thus, for example,

$$\lambda_{11}(b_{11}, b_{12}, \ldots, b_{1(sn)}) = (\lambda_{11}b_{11}, \lambda_{11}b_{12}, \ldots, \lambda_{11}b_{1(sn)}),$$

where each entry on the right is interpreted modulo p.

We thus have shown that a code corresponding to a single generator function g(t) and the subfield F of order p^r may be viewed as a code corresponding to r generator functions $g_1(t), \ldots, g_r(t)$ and the prime field of order p. We turn now to Remark 2b.

We are concerned here with showing that it is possible to transform a code that is known to have good distance properties into canonic form and still preserve those distance properties.

We suppose that the n-tuple corresponding to g(t) has been transformed, as discussed in Remark 2a, into a set of r sn-tuples with entries in G. F.[p]. We again designate these sn-tuples by $\hat{g}_1(t), \ldots, \hat{g}_r(t)$.

Clearly, we do not alter the code generated by $\hat{g}_1(t), \ldots, \hat{g}_r(t)$ and the prime field G. F. [p] by

(1) replacing any generator $\hat{g}_i(t)$ by $\lambda\hat{g}_i(t)$ if λ is a nonzero element of G. F.[p];

(2) replacing $\hat{g}_i(t)$ by $\hat{g}_i(t) + \lambda\hat{g}_j(t)$, where λ is any element of G. F.[p];

(3) interchanging $\hat{g}_i(t)$ with $\hat{g}_j(t)$ for any i, j $1 \le i$, $j \le r$.

These operations are analagous to the "elementary row operations" of matrix theory by means of which it is possible to reduce any matrix into a row-reduced echelon form without affecting the space spanned by the matrix.[7]

If, in addition to these three operations, we allow column permutations of the generator array, then it is possible to derive from the original set of generators $\hat{g}_1(t), \ldots, \hat{g}_r(t)$ a new set of generators, say $p_1(t), \ldots, p_r(t)$, which have the forms

100	...	0	c_{11}	...	$c_{1(sn-r)}$
010	...	0	c_{21}	...	$c_{2(sn-r)}$
$\vdots$					
000	...	1	c_{r1}	...	$c_{r(sn-r)}$

respectively, where the c_{ij} belong to the prime field of order p.

In general, column permutations of the generator array will alter the code. Thus the code generated by $\hat{g}_1(t), \ldots, \hat{g}_r(t)$ may differ from the code generated by $p_1(t), \ldots, p_r(t)$. However, if the set of allowable column permutations is suitably restricted, the two codes will have the same metric structure.

It is important to keep in mind here that we are still measuring distance according to the rules of the Galois field of order p^s, even though all computations are carried out

modulo p. Thus to calculate the Hamming distance between two vectors it is necessary to compare their entries in successive blocks of length s. Each such block comparison can contribute only 1 to the cumulative Hamming distance between the two vectors, regardless of the actual number of discrepancies in the block which are in excess of 1.

Consider, for example, the two code vectors

$$d_1, d_2, \ldots, d_{sn}$$

$$f_1, f_2, \ldots, f_{sn}.$$

There are n blocks of length s to be compared. The first comparison involves a check of

$$d_1, \ldots, d_s$$

$$f_1, \ldots, f_s.$$

If there is disagreement in one or more places, these two blocks are distance 1 apart.

Compare, next, the s entries

$$d_{s+1}, \ldots, d_{2s}$$

$$f_{s+1}, \ldots, f_{2s} \quad .$$

Again, disagreement in one or more places adds 1 to the cumulative Hamming distance between the two code vectors. Clearly, the maximum distance between the two code vectors is equal to n.

It should be clear that any two columns that appear in the same block of length s can be permuted without changing the distance between the two code words.

Thus, for example, the distance between the vectors

$$d_1, d_2, \ldots, d_s, d_{s+1}, \ldots, d_{sn}$$

$$f_1, f_2, \ldots, f_s, f_{s+1}, \ldots, f_{sn}$$

is equal to the distance between the two vectors

$$d_s, d_2, \ldots, d_{s-1}, d_1, d_{s+1}, \ldots, d_{sn}$$

$$f_s, f_2, \ldots, f_{s-1}, f_1, f_{s+1}, \ldots, f_{sn}.$$

This is not true if columns from distinct blocks are interchanged. Thus, in general, the distance between the last two vectors given above is not equal to the distance between the vectors

$$d_{s+1}, \ d_2, \ \ldots, \ d_s, \ d_1, \ d_{s+2}, \ \ldots, \ d_{sn}$$

$$f_{s+1}, \ f_2, \ \ldots, \ f_s, \ f_1, \ f_{s+2}, \ \ldots, \ f_{sn}.$$

Consider now the first s entries of the generator functions $\hat{g}_1(t), \ \ldots, \ \hat{g}_r(t)$, namely,

$$b_{11}, \ \ldots, \ b_{1s}$$

$$b_{21}, \ \ldots, \ b_{2s}$$

$$b_{r1}, \ \ldots, \ b_{rs}.$$

In accordance with this discussion, we can permute the columns of this matrix of entries without affecting the metric structure of the portion of the code consisting of elements of the form

$$\lambda_1 \hat{g}_1(t) + \ldots + \lambda_r \hat{g}_r(t),$$

where the λ_i are chosen from the prime field of order p.

The total code, however, involves translates of the generator functions. Thus the elements that appear in the first block of generator entries, in the process of forming the code, will interact with elements that appear in blocks numbered n_o+1, $2n_o+1$, $\ldots$, $(k-1)n_o+1$, and, in fact, only with these blocks. It follows that the metric structure of the code will be preserved if any permutations of the columns of block 1 are paralleled in these blocks. Thus if the first column of block 1 is interchanged with the third column of block 1, then the first column of block n_o+1 should be interchanged with the third column of block n_o+1, and similarly for blocks numbered $2n_o+1$, $\ldots$, $(k-1)n_o+1$.

The manipulations discussed above permit us to derive the set of generators $p_1(t)$, $\ldots$, $p_r(t)$ whose first r entries are in diagonal form, as shown earlier.

Finally, we note that if $h_1(t)$, $\ldots$, $h_r(t)$ are any paths belonging to the code subsets containing $p_1(t)$, $\ldots$, $p_r(t)$, respectively, then the code generated by $h_1(t)$, $\ldots$, $h_r(t)$ is identical to the code generated by $p_1(t)$, $\ldots$, $p_r(t)$.[8] Using this fact, we can extract a set of generators for the code which are in canonic form. That is to say, the generators may be written in the form

where the symbol x denotes some entry from the prime field. (Note that entries in different positions will, in general, have different values.)

We conclude this report with an example designed to illustrate Remark 2. We consider a generator with entries drawn from G. F.[16] (see Peterson[5]) and assume that $n = 6$, $n_o = 2$, and $m = 4$. As noted earlier, G. F.[16] may be represented as the field of polynomials over G. F.[2] with multiplication performed modulo the polynomial $1 + x + x^4$. The subfield E of order 4 consists of the elements 1, $x + x^2$, $1 + x + x^2$, 0. Clearly E, considered a vector space of dimension 2 over the binary field, is spanned by the two basis vectors 1 and $x + x^2$. Let us assume a generator function $g(t)$ of the form

$$1 + x + x^3, \; x + x^2 + x^3, \; 1 + x^2 + x^3, \; x + x^3, \; x^2 + x^3, \; 1 + x^2.$$

Consider the pair of functions $(1)g(t)$ and $(x+x^2)g(t)$:

$$1 + x + x^3, \; x + x^2 + x^3, \; 1 + x^2 + x^3, \; x + x^3, \; x^2 + x^3, \; 1 + x^2$$

$$1 + x + x^2 + x^3, \; x, \; x^3, \; 1 + x^3, \; x + x^2 + x^3, \; 1 + x^2 + x^3.$$

Denoting these two 6-tuples by $g_1(t)$ and $g_2(t)$, respectively, we may write the corresponding 24-tuples $\hat{g}_1(t)$ and $\hat{g}_2(t)$ as

110101111011010100111010

111101000001100101111011.

Replacing $\hat{g}_2(t)$ by $\hat{g}_1(t) + \hat{g}_2(t)$, we get the pair

110101111011010100111010

001000111010110001000001.

We can put the first two entries of this array into diagonal form by interchanging columns 2 and 3. Correspondingly, in order to preserve the distance properties of the tree code generated by these two vectors, we must further interchange column 10 with 11 and column 18 with column 19. We denote the resulting pair of vectors

101101111010101011010

010000111100110000100001

as $p_1(t)$ and $p_2(t)$, respectively.

The code generated by $p_1(t)$ and $p_2(t)$ consists of the 2^6 vectors

$$\sum_{i=0}^{2} \lambda_{1i} p_1(t-i8) + \sum_{i=0}^{2} \lambda_{2i} p_2(t-i8),$$

where the λ_{ji} are elements of the binary field.

The canonic generators for this code are the two vectors

$$h_1(t) = p_1(t) + p_1(t-8) + p_2(t-8) + p_2(t-16)$$

$$h_2(t) = p_2(t) + p_1(t-8) + p_2(t-8).$$

Note that $h_1(t)$ belongs to the path subset containing $p_1(t)$, and $h_2(t)$ belongs to the path subset containing $p_2(t)$. Computing $h_1(t)$ and $h_2(t)$ explicitly, we find that they are equal to the pair of 24-tuples

101101110010000100000000

010000110011100000111000,

respectively.

The important thing to note about this array is that it is of the form

The author wishes to acknowledge his indebtedness to Professor John M. Wozencraft for stimulating this research. The basic comparison of the number of low-weight code words and essentially distinct codes, which was used to prove the theorem, is a generalization of his argument for binary fields.[9] The observation that codes may be transformed into canonic form without destroying their metric structure is also due to him.

H. Dym

References

1. W. W. Peterson, Error-Correcting Codes (The M. I. T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961), p. 51.

2. $p(t)$ is only of interest for $t = 1, \ldots, n$; the n-tuple corresponding to $p(t)$ is, of course, $(p(1), p(2), \ldots, p(n))$.

3. It actually would have been sufficient to require only that one of the first n_0 entries of $g(t)$ be nonzero in order to guarantee the linear independence of $g(t)$ and its translates.

4. R. D. Carmichael, Introduction to the Theory of Groups of Finite Order (Dover Publications, Inc., New York, 1956), p. 260.

5. This example, for ease of reference, was chosen to coincide notationally with a discussion of G. F.[16] which appears in Peterson, op. cit., p. 100.

6. This may require a very long time, however, especially when n and k are large.

7. K. Hoffman and R. Kunze, Linear Algebra (Prentice-Hall, Inc., Englewood Cliffs, N. J., 1961).

8. Recall, for example, that $h_1(t)$ belongs to the same subset as $p_1(t)$ if and only if $h_1(t) = p_1(t)$ for $t = 1, 2, \ldots, sn_0$.

9. J. M. Wozencraft and B. Reiffen, Sequential Decoding (The M. I. T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961), pp. 51-54.

C. MOMENTS OF THE SEQUENTIAL DECODING COMPUTATION

Recent investigations of a sequential decoding algorithm[1] for the memoryless binary erasure channel provide results with implications for the behavior of sequential decoding on the general, memoryless channel.

The behavior of the moments of the sequential decoding computation as a function of rate on the erasure channel has been determined. The nth moment grows exponentially with constraint length for rates $R > R_n$, $n = 1, 2, 3, \ldots$. The rates $\{R_n\}$ form a mono-tonically decreasing sequence with an interesting geometrical interpretation. Plot the exponent, $E(R)$, on the "sphere-packed" probability of error versus R (see Fig. XV-4).
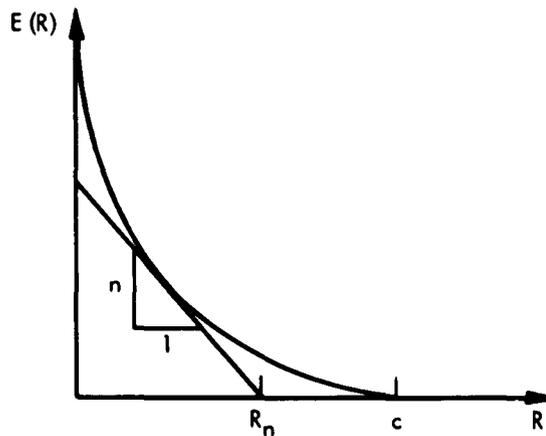
Fig. XV-4. Rate construction.

Draw a line with slope equal to $-n$ tangent to $E(R)$. Then the rate-axis intercept of this straight line is the rate R_n. This is a very natural extension of the geometrical inter-pretation of $R_{comp} \equiv R_1$. Because the rates $\{R_n\}$ have such a natural interpretation and because sequential decoding on the erasure channel exhibits the fundamental features that are exhibited on more general channels, we are inclined to believe that these results also apply to the general, memoryless channel. Preliminary investigations indicate that this is true.

The behavior of the rates $\{R_n\}$ implies something about the character of the

distribution of computation, $P_R[x \geqslant N]$. In particular, if we assume that the fractional rates can be obtained by the same construction technique as the integral rates, then the Pareto distribution

$$P_R[x \geqslant N] = \frac{A}{N^\beta}, \ N \text{ large and } \beta \text{ such that } R = R_\beta$$

provides the correct moment behavior. This problem, as well as those mentioned earlier, is still under study.

<div align="right">J. E. Savage</div>

References

1. For a description of the algorithm employed, see J. E. Savage, Sequential decoding for an erasure channel with memory, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M.I.T., April 15, 1963, pp. 149-154.

XVI. SPEECH COMMUNICATION*

Prof. K. N. Stevens
Prof. M. Halle
Prof. J. B. Dennis
Prof. J. M. Heinz

Dr. A. S. House
Jane B. Arnold
W. L. Henke

A. P. Paul
S. S. Reisman
T. E. Taylor
E. C. Whitman

A. REAL-TIME SPEECH SPECTRUM ANALYZER

As the M. I. T. Speech Synthesizer (DAVO) will soon be controlled by the TX-0 digital computer, it is desirable to have a speech spectrum analyzer that can make use of a real-time speech input so that synthesized utterances may be subjected to immediate analysis. A general system design for a real-time speech spectrum analyzer has been formulated, and a tentative design for some of the circuitry has been completed.

Figure XVI-1 is a function block diagram of the proposed system. A power amplifier is used to drive 36 bandpass filters that have the same bandpass characteristics as the
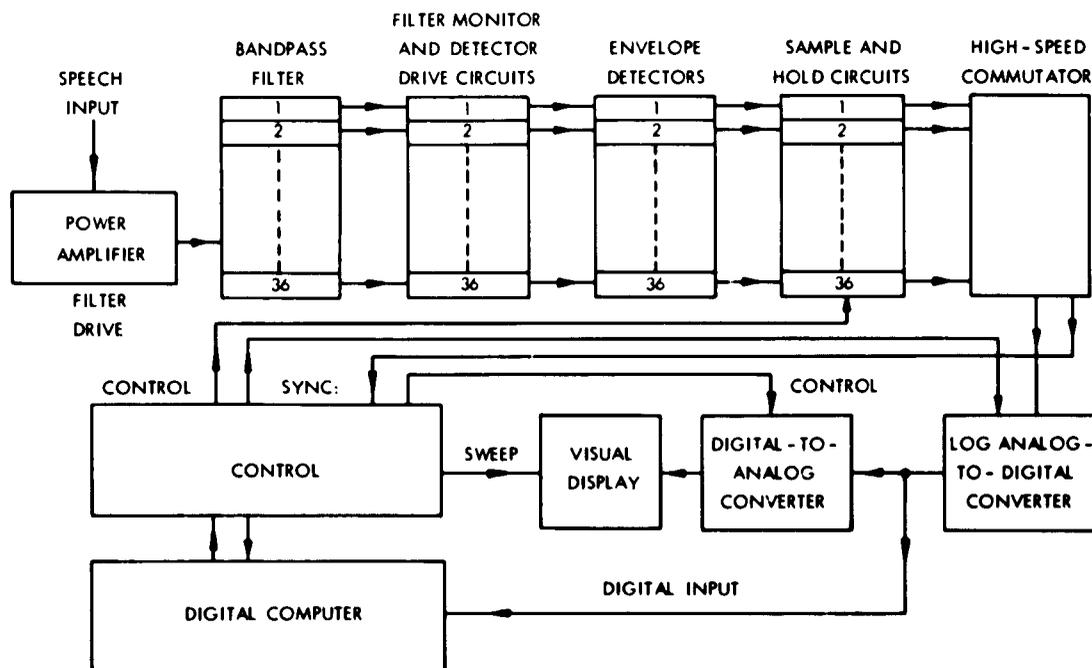

Fig. XVI-1. Function diagram of the real-time speech analyzer.

filters now in use for speech analysis. The low-level filter output is monitored and amplified to a level that is suitable for driving the envelope detector. The detector output is sampled on command and stored in a holding circuit. The holding circuit outputs are serially scanned by a high-speed commutator, converted to digital form and used as input data to a digital computer.

The bandpass filters, filter drive, filter-output monitor and envelope-detector circuits were developed in detail to determine the best circuit configurations that are consistent with the requirements of cost, reliability, and simplicity. More detailed information on these circuits can be found in the author's thesis.[1]

<div align="right">P. S. Marchese</div>

References

1. P. S. Marchese, A High Speed Speech Analyzer for Use with a Digital Computer, S. B. Thesis, Department of Electrical Engineering, M. I. T., June 1963.

XVII. MECHANICAL TRANSLATION[*]

V. H. Yngve	Muriel S. Kannel	T. More, Jr.
Carol M. Bosche	E. S. Klima	W. K. Percival
Elinor K. Charney	E. S. Lowry	A. C. Satterthwait
Ann Congleton	J. D. McCawley	K. G. Sellin
J. L. Darlington		J. S. Wright

A. LOGICAL EVALUATION OF ARGUMENTS STATED IN "FORMAT Q"

The COMIT program for logical translation and evaluation (Quarterly Progress Reports No. 68 (pages 174-175) and No. 69 (pages 165-168)) has been developed to the point at which one may submit for evaluation an entire argument written in a quasi-logical notation, "format Q." The program translates the argument into a strictly logical functional calculus notation, "format L," and then proceeds to test its validity by using the Davis-Putnam proof-procedure algorithim.[1] The following excerpts from the machine output resulting from the translation and evaluation of a sample argument may be presented and briefly discussed. The sample argument, as it originally occurred in a logic textbook,[2] is

> "Whoever belongs to the Country Club is wealthier than any member of
> the Elks Lodge. Not all who belong to the Country Club are wealthier
> than all who do not belong. Therefore not everyone belongs either to
> the Country Club or the Elks Lodge."

This argument was translated by hand into the following "format Q" representation; in this form it was submitted to the machine, which then proceeded to translate it into "format L," test it, and find it to be valid, in the time of 0.7 minute, exclusive of compilation.

> THE INPUT ARGUMENT IS ALL + X/A + SUCH + THAT + X/A + BELONGS +
> TO + THE + COUNTRY + CLUB + IS + WEALTHIER + THAN + ALL + X/B +
> SUCH + THAT + X/B + BELONGS + TO + THE + ELKS + LODGE + . + SOME +
> X/C + SUCH + THAT + X/C + BELONGS + TO + THE + COUNTRY + CLUB +
> IS + NOT + WEALTHIER + THAN + SOME + X/D + SUCH + THAT + X/D +
> BELONGS + NOT + TO + THE + COUNTRY + CLUB + . + THEREFORE + SOME
> + X/E + SUCH + THAT + X/E + BELONGS + NOT + TO + THE + COUNTRY +
> CLUB + IS + AN + X/E + SUCH + THAT + X/E + BELONGS + NOT + TO + THE
> + ELKS + LODGE + .

The sample argument consists of three sentences, the first two of which are premises and the third of which is the conclusion, since its first word is 'therefore'. The program proceeds to parse each sentence individually, in accordance with the grammar

described in Quarterly Progress Report No. 69 (pages 165-168). In the course of parsing, each word or sequence of words constituting what logicians call a "predicate," either simple or relational, is labelled with 'P' and given a numerical subscript. In our example, 'belongs to the Country Club' is labelled with 'P/.37', 'belongs to the Elks Lodge' with 'P/.38', and 'is wealthier than' with 'P/.225'. For the negative predicates, like 'belongs not to the Country Club', the subscript '/NOT' is added to the appropriate 'P'.

The structure of each parsed sentence corresponds to an equivalent structure in format L. The program contains a list of these equivalences and uses them to translate each sentence into format L. For all but the most simple sentences, the translation into format L involves several applications of these structural equivalences, since the first application usually results in a formula containing parts that are not yet in format L. These parts are then translated in turn into format L, and the results of these translations may themselves contain parts that need further translation. This translative "loop" is repeatedly executed until each sentence is entirely in format L. Any resulting formula that is not already in prenex normal form, with all the quantifiers on the left, is run through a subroutine that puts it into this form.

The three prenex formulae, corresponding to the three sentences of the argument, were printed out by the machine as follows:

(1) THE PRENEX NORMAL FORM IS Q/A, ALL + Q/B, ALL + *(+ *(+ *(+ P/.37, A + *) + AND/C + *(+ P/.38, B + *) + *) + IMPLIES/C + *(+ P/.225 + *(+ X/A + , + X/B + *) + *) + *) +

(2) THE PRENEX NORMAL FORM IS Q/SOME, C + Q/SOME, D + *(+ *(+ *(+ P/.37, C + *) + AND/C + *(+ P/.37, NOT, D + *) + *) + AND/C + *(+ P/.225, NOT + *(+ X/C + , + X/D + *) + *) + *) +

(3) THE PRENEX NORMAL FORM IS Q/SOME, E + *(+ *(+ P/. 37, NOT, E + *) + AND/C + *(+ P/.38, NOT, E + *) + *) +

These formulae are then combined into a single formula, of implicational form, in which the conjunction of the premises is taken to imply the conclusion. The single formula is then put into prenex normal form and undergoes some added changes in format. All of the information contained in the subscripts, numerical or otherwise, is incorporated into the symbols themselves and the subscripts are eliminated. The principal reason for these changes in format is that the subsequent proof-procedure program, which was actually written before the translation program, was written without using any subscripts. This conversion of format has been greatly facilitated by the recent addition to the COMIT system of a provision for elevating any subscript on a symbol into a symbol in its own right (it may also work the other way around – a symbol may be turned into a subscript). The resulting for.nula, representing the input argument,

was printed out by the machine as follows:

ENTERING PROOF PROCEDURE THE FORMULA IS (EA)(EB)(AC)(AD)(EE)(((((P37A)
AND(P38B))IMPLIES(P225AB))AND(((P37C)AND(NOT(P37D)))AND(NOT(P225CD))))
IMPLIES((NOT(P37E))AND(NOT(P38E))))

The proof-procedure program, based on the Davis-Putnam algorithm, operates by reductio ad absurdum, that is, by negating the formula and attempting to derive a contradiction. For the sample argument, the negated formula consists of a sequence of quantifiers,

'(AA)(AB)(EC)(ED)(AE)',

followed by a matrix in conjunctive normal form,

'P37C AND NP37D AND NP225CD AND NP37A NP38B P225AB AND P37E
P38E AND'.

The existentially quantified variables, 'C' and 'D', are replaced in the matrix by 'PAB' and 'QAB', respectively, which are distinct functions of 'A' and 'B', the universally quantified variables that precede 'C' and 'D' in the sequence of quantifiers. This gives the matrix

'P37PAB AND NP37QAB AND NP225PABQAB AND NP37A NP38B P225AB
AND P37E P38E AND'.

In evaluating the sample argument, the program generated a sequence of $3^3 = 27$ "quantifier-free lines" on the basis of this matrix, by substituting the terms 'A', 'PAA', and 'QAA' for the variables 'A', 'B', and 'E' in all possible combinations. These 27 lines were found to contain a contradiction; thus the original formula is valid.

In the immediate future, it is hoped that the program described can be improved in some or all of the following ways:

(1) By mechanizing the translation from ordinary language into format Q, or at least from a restricted ordinary language into format Q.

(2) By expanding the list of quantifier-words, at present restricted to 'all', 'some', 'no', 'only', and 'the', so as to allow for numerical propositions. The program already permits 'at most n', where 'n' is a whole number equal to or less than 20, to be used as a quantifying expression. We next plan to program 'at least n' and 'exactly n', the latter of which will be treated as the conjunction of 'at most n' and 'at least n'.

(3) By expanding the grammar so as to admit a greater variety of sentence-types, at present restricted to sentences in which two "NPs" (noun phrases) are connected by a form of the verb 'to be', or by a "binary relational predicate," such as 'is wealthier than'. It eventually will be desired to handle relational predicates of greater degree, such as the ternary predicate illustrated by the construction 'A gives B to C'.

<div align="right">J. L. Darlington</div>

References

1. M. Davis and H. Putnam, A computing procedure for quantification theory, J. Assoc. Computing Machinery 7, 201-215 (1960).

2. I. M. Copi, Symbolic Logic (Macmillan and Company, New York, 1958), p. 140, no. 2.

B. CONSONANT MUTATION IN FINNISH

1. Introduction

In all Finnish words except those to be noted in section 3, a noninitial stop at the beginning of a short closed syllable undergoes various changes that depend on the surrounding segments, and thus produce alternations when an affix such as the genitive ending -n is added which closes a syllable that had hitherto been open. The changes are as follows:

(a) single /p, t, k/ become /v, d, zero/, respectively, except that /k/ becomes /j/ after /h/ or between a liquid and /i/ or /e/, /k/ becomes /v/ between two /u/'s or /y/'s,[1] and there is no mutation if the stop is preceded by /s/ (also sometimes if preceded by /h/) or in the combination /tk/.

	'help'	'worm'	'river'		'straw'	'calf (of leg)'	'dress'
Nominative	apu	mato	joki	but	olki	pohjeh	puku
Genitive	avun	madon	joen		oljen	pohkeen	puvun

(b) A geminate stop is shortened:

	'priest'	'carpet'	'flower'
Nominative	pappi	matto	kukka
Genitive	papin	maton	kukan

(c) Nasal plus single stop becomes geminate nasal:

	'coast'	'warmth'	'thread'
Nominative	ranta	lämpö	lanka
Genitive	rannan	lämmön	laŋŋan (written langan)

(d) Liquid plus /t/ becomes geminate liquid:

	'evening'	'stream'
Nominative	ilta	virta
Genitive	illan	virran

2. Mutation Rules

I shall endeavor here to state a maximally simple set of rules to generate correctly all forms with consonant mutation. Note first that if the above-given statement of the

facts is translated directly into distinctive-feature terms, the resulting rules would be extremely cumbersome. A much simpler set of rules can be obtained by describing the process in a somewhat different fashion, namely, by saying that <u>all</u> stops undergo some particular change in the environment[2]_____ C $\left\{{C \atop \#}\right\}$ and that the resulting forms are converted into the correct forms by other rules that perform assimilations or deletions.

It is necessary to assume the following set of morphophonemically distinct consonants for the base forms of Finnish morphemes: p, t, k, v, s, h, m, n, l, r.[3] Note that the feature of voicing is nondistinctive, that is, predictable: the features of obstruence, continuence, nasality, and those features relating to place of articulation (/p, k/ are grave and /t/ nongrave; /k/ is compact and /p, t/ noncompact) are sufficient to distinguish between these segments. The segments that are voiced are the vowels, the resonants (liquids and nasals), and /v/. Voicing can thus be predicted by the following rules:

$$[\quad] \rightarrow [- \text{voice}]$$

$$\left\{{\begin{matrix} [- \text{obs}] \\ \begin{bmatrix} + \text{obs} \\ + \text{cnt} \\ + \text{grv} \\ - \text{cmp} \end{bmatrix} \end{matrix}}\right\} \rightarrow [+ \text{voice}]$$

that is, everything becomes voiceless except nonobstruents (vowels, liquids, and nasals) and /v/, which become voiced.

Since the result of consonant mutation is a voiced segment except in only those cases in which a segment is deleted entirely, I propose having the consonant mutation rule make <u>all</u> stops voiced in the given environment and then have a set of assimilation and deletion rules that apply to voiced stops. This will necessitate, of course, that the rules that predict voicing occur earlier in the grammar than the consonant-mutation rules. I thus state the rule:

1. $[- \text{cnt}] \rightarrow [+ \text{voice}]$ in env _____ V C $\left\{{C \atop \#}\right\}$.

This, among other things, converts the geminate stops <u>pp</u>, <u>tt</u>, <u>kk</u> into <u>pb</u>, <u>td</u>, <u>kg</u>. Since geminate stops become single stops under consonant mutation, a rule will be needed to delete a voiced stop preceded by a homorganic voiceless stop:

2. $\begin{bmatrix} + \text{obs} \\ + \text{voice} \\ \alpha \text{ grv} \end{bmatrix} \rightarrow \emptyset$ in env $\begin{bmatrix} + \text{obs} \\ - \text{cnt} \\ \alpha \text{ grv} \end{bmatrix}$ _____ .

The specification of $[\alpha \text{ grv}]$[4] ensures that the rule will apply only to homorganic-stop sequences. Note that /tk/, the only nonhomorganic-stop sequence, does not undergo consonant mutation.

The combinations sp, st, sk, and tk are unaffected by consonant mutation. That means that the sb, sd, sg, and tg into which rule 1 would convert them must be made voiceless again. This is accomplished by the following rule:

3. $[+ \text{obs}] \rightarrow [- \text{voice}]$ in env $\begin{bmatrix} + \text{obs} \\ - \text{grv} \end{bmatrix}$ _____ .

There next follows a set of assimilation rules:

4. $\begin{bmatrix} + \text{obs} \\ + \text{voice} \end{bmatrix} \rightarrow \begin{bmatrix} + \text{nas} \\ - \text{obs} \end{bmatrix}$ in env $[+ \text{nas}]$ _____ .

5.
$$\begin{bmatrix} + \text{obs} \\ + \text{voice} \\ + \text{grv} \end{bmatrix} \rightarrow \left\{ \begin{array}{l} [- \text{cmp}] \quad \text{in env} \quad \begin{bmatrix} - \text{cns} \\ + \text{flt} \\ + \text{dif} \end{bmatrix} \underline{\quad} \begin{bmatrix} - \text{cns} \\ + \text{flt} \\ + \text{dif} \end{bmatrix} \\[2em] \begin{bmatrix} - \text{cns} \\ - \text{obs} \\ - \text{grv} \end{bmatrix} \quad \text{in env} \quad \left\{ \begin{array}{l} \begin{bmatrix} + \text{cns} \\ + \text{voc} \end{bmatrix} \underline{\quad} \begin{bmatrix} - \text{cns} \\ - \text{flt} \\ - \text{cmp} \end{bmatrix} \\[1.5em] \begin{bmatrix} - \text{cns} \\ - \text{voc} \\ + \text{cmp} \end{bmatrix} \underline{\quad} \end{array} \right\} \end{array} \right\} .$$

6. $\begin{bmatrix} + \text{obs} \\ + \text{voice} \\ - \text{grv} \end{bmatrix} \rightarrow \begin{bmatrix} - \text{obs} \\ \alpha \text{ cnt} \end{bmatrix}$ in env $\begin{bmatrix} + \text{cns} \\ + \text{voc} \\ \alpha \text{ cnt} \end{bmatrix}$ _____ .

Rule 4 converts /b, d, g/ into nasals when they are preceded by nasals, and thus has the total effect of creating geminate nasals out of nasal-stop sequences. The first part of rule 5 changes /g/ into /b/ (which later becomes /v/) when it occurs between two /u/'s or /y/'s, and the second part converts it into a /j/ when it occurs in the environment Liquid_____ $\left\{ \begin{matrix} i \\ e \end{matrix} \right\}$ or /h/_____ . Rule 6 assimilates /d/ to a preceding liquid.

It remains only to delete all remaining /g/'s (in accordance with the fact that consonant mutation converts /k/ into zero in all environments except those discussed above) and to convert all remaining /b/'s into /v/'s. This is accomplished by the rules:

7. $\begin{bmatrix} + \text{obs} \\ + \text{cmp} \\ + \text{voice} \end{bmatrix} \rightarrow \emptyset$

8. $\begin{bmatrix} + \text{obs} \\ + \text{grv} \\ + \text{voice} \end{bmatrix} \rightarrow [+ \text{cnt}]$.

Rules 1-8 are part of a larger set of rules that have been tested by means of a computer program (written in COMIT) which executes the rules on the base forms of a given set of words and prints out the results both in the form of a matrix of distinctive-feature specifications and in a phonemic orthography. The results produced by the machine are

correct for all words that I have tried thus far.

3. Environment in which Consonant Mutation Operates

I stated above that consonant mutation occurs at the beginning of a short closed syllable and gave the formula _____ V C $\left\{\begin{matrix} C \\ \# \end{matrix}\right\}$. I must now make more precise just what is to be understood by "short closed syllable," which will amount to stating what the C's are to include. Consonant mutation occurs before syllables closed by a glide; for example, the partitive plural of silakka 'herring' is silakojta (written silakoita in standard orthography). The features of consonantalness and vocalicness distinguish between true consonants, liquids, glides, and vowels as follows:

	true consonants	liquids	glides	vowels
cns	+	+	–	–
voc	–	+	–	+

The first C of the formula given thus can be any segment that is either [+ cns] or [– voc], that is, anything except a vowel. There is dialect variation as to whether the second C may be a glide or whether it can only be a liquid or true consonant. In the former type of dialect the illative plural of silakka is silakojhin, and the rule has $\left\{\begin{matrix} [+ \text{cns}] \\ [- \text{voc}] \end{matrix}\right\}$ for the second C; in the latter type of dialect the illative plural is silakkojhin, and the rule has [+ cns]. The reformulation of rule 1 which is given below is for the latter type of dialect. Both silakojhin and silakkojhin seem to be equally frequent in educated Finnish speech. Finally, consonant mutation does not apply to an initial stop; for example, pappi does not become *vappi. This means that the rule must require that the stop in question be preceded by at least one segment. The full form of the rule, thus, is

1. [– cnt] → [+ voice] in env [   ] _____ V $\left\{\begin{matrix} [+ \text{cns}] \\ [- \text{voc}] \end{matrix}\right\}\left\{\begin{matrix} [+ \text{cns}] \\ \# \end{matrix}\right\}$

With respect to my assertion that consonant mutation occurs before short syllables closed by /j/, it may be objected that there are words in which a nonmutated consonant occurs before a VjC sequence; for example, the partitive plural of hammas 'tooth' is hampaita and not *hammaita. However, all of the words in which this happens have underlying forms in which there is a consonant between the vowel and the /j/. For example, hampaita is derived from the underlying form hampas + i + ta. Thus it suffices to assume that the rule for deleting the final consonant in these words occurs later in the grammar than the consonant-mutation rule. Then, at the time the consonant-mutation rule applies, the stops in question will not be at the beginning of a short closed syllable, so that the rule will have no effect.

4. Words That Do Not Undergo Consonant Mutation

There are a large number of proper names and incompletely assimilated loanwords that do not undergo consonant mutation. These words are also peculiar in many other respects and simply will have to be treated as a different class of words, to which certain of the morphophonemic rules will not apply.

However, there are also two situations in which native Finnish words do not undergo consonant mutation. One is words in which the closed syllable is produced by a possessive affix, for example, tupa 'hut', tupamme 'our hut' (not *tuvamme). The other is words with the suffix -nen~-s-. These words have an -s- throughout their paradigms, except in the nominative singular in which it is replaced by -nen. The -nen~-s- must be treated as a separate morpheme; it often functions as either an adjective-forming or a diminutive suffix, although there are many words such as hevonen 'horse' in which it does not have such a function. In words of this type there is no consonant mutation in the partitive, even though a short closed syllable is formed.

	'sparrow'	'preceding'
Nominative	varpunen	entinen
Genitive	varpusen	entisen
Partitive	varpusta (not *varvusta)	entistä (not *ennistä)

This amounts to saying that, for the purposes of the application of consonant mutation, possessive suffixes and -s- do not count as part of the word or, what is the same thing, for the purposes of consonant mutation the words in question are split into two parts, and the rule applies to these parts separately rather than to the entire word. This is, of course, the familiar phenomenon of "juncture," about whose precise nature there has been much controversy, but which, as the word is used by all authors with whom I am familiar, consists in the splitting of an utterance into chunks that serve as the domains on which the various morphophonemic processes operate. Authors such as Trager,[5] Hill,[6] and Joos[7] speak of juncture as an independent phonetic entity; but the places in which they observe this alleged element are simply places in which a segment in the middle of an utterance behaves in some respects as if it were at the beginning or end of the utterance, so that the phenomenon still can be regarded as one of the splitting of an utterance into chunks.[8] Thus, by marking the splits in the words in question by /, tupamme and entistä are represented as tupa/mme and enti/s + tä (+ denotes morpheme boundary).

Nouns in -nen~-s- have another peculiarity, namely the fact that the obstruent-deletion rule referred to earlier does not apply to them. The rule in question deletes an intervocalic obstruent if it occurs later than the second syllable and is preceded by a single vowel.[9] The rule is involved in the generation of the oblique case forms of nouns with stems ending in a single obstruent. For example, the genitive case hampaan of the noun hammas is generated as follows:

underlying form	hampas + e + n
obstruent deletion	hampa + e + n
vowel assimilation	hampa + a + n

However, <u>entisen</u> does not become *<u>entiin</u>. The assumption of a juncture / before the affix -<u>nen</u>~-<u>s</u> would give an underlying form enti / s + e + n for <u>entisen</u>. For the purposes of the obstruent-deletion rule the word again would be broken up into two chunks, <u>enti</u> and <u>s</u> + <u>e</u> + <u>n</u>, and the obstruent-deletion rule would not apply to either. Representing -<u>nen</u>~-<u>s</u> nouns with a / before that affix thus simultaneously accounts for the fact that -<u>nen</u>~-<u>s</u> nouns are the only words in the native vocabulary (outside of possessive forms) which do not undergo consonant mutation and the fact that they are also the only words that have intervocalic /s/'s beyond the second syllable.

J. D. McCawley

References

1. I write my examples in a slightly modified version of standard Finnish orthography. I denote final "aspiration" by <u>h</u> and write <u>j</u> for phonetic /j/ even where the standard orthography writes <u>i</u>. <u>y</u> and <u>ä</u> denote the sounds represented in the International Phonetic Alphabet by <u>ü</u> and <u>æ</u>.

2. # means word boundary. The actual environment is somewhat more complicated. I state exactly what it is in section 3.

3. <u>d</u> and <u>ŋ</u>, which are regarded as separate phonemes by most American linguists, only arise through constant mutation or assimilation and thus do not have to be recognized as separate morphophonemes.

4. Greek letters α, β, ... denote variables that assume the values + and -. Variables are useful in stating identities, as in rule 2, and in formulating assimilation and dissimilation rules. For example, $[+ \text{cns}] \rightarrow [\alpha \text{ voice}]$ in env_____ $\begin{bmatrix} + \text{ cns} \\ \alpha \text{ voice} \end{bmatrix}$ would be a rule for the regressive assimilation of voicing in consonant sequences.

5. G. L. Trager and H. L. Smith, Jr., <u>Outline of English Structure</u> (University of Oklahoma Press, Norman, 1951).

6. A. A. Hill, <u>Introduction to Linguistic Structures</u> (Harcourt, Brace and Company, New York, 1958).

7. M. Joos, <u>English Language and Linguistics</u> (Institute for Experimental Phonetics, Beograd, Yugoslavia, 1958).

8. It will be necessary, of course, to have several different junctural elements, since some morphological processes apply to larger domains than do others.

9. The present form of this rule was suggested to me by R. P. Kiparski. An earlier and incorrect version of the rule is given in J. D. McCawley, Finnish noun morphology, Quarterly Progress Report No. 68, Research Laboratory of Electronics, M.I.T., January 15, 1963, pp. 180-186.

XVIII. LINGUISTICS[*]

Prof. R. Jakobson
Prof. A. N. Chomsky
Prof. M. Halle
Prof. Y. Isami
Dr. G. H. Matthews
Dr. Paula Menyuk

Dr. K. Wu
T. G. Bever
S. K. Ghosh
Barbara C. Hall
J. J. Katz
R. P. V. Kiparsky
S.-Y. Kuroda

D. T. Langendoen
T. M. Lightner
P. M. Postal
C. B. Qualls
J. J. Viertel
A. M. Zwicky, Jr.

A. SEMANTIC INTERPRETATION OF IDIOMS AND SENTENCES CONTAINING THEM

1. Two Types of Idioms

In the present report we examine the semantics of idioms in natural languages. We show how a class of idioms may be treated in terms of the recently developed conception of the semantic component of a linguistic description.[1-4] Familiarity with this conception is assumed throughout.

The essential feature of an idiom is that its full meaning, and more generally the meaning of any sentence containing an idiomatic stretch, is not a compositional function of the meanings of the idiom's elementary grammatical parts. For example, the meaning of the idiom kicked the bucket cannot be regarded as a compositional function of the meanings kick, ed, the, bucket, regardless of the syntactic structure attributed to kicked the bucket by the structural descriptions of the sentences in which it appears. Hence the projection rules that a semantic theory provides to obtain the meaning of compound expressions and sentences as a compositional function of the meanings of their elementary parts cannot obtain the idiomatic meaning of an idiomatic stretch from the meanings of the syntactically atomic parts of that stretch. Therefore, the fact that no projection rules at all are employed in obtaining semantic interpretations for whole idiomatic stretches is the formal representation that a semantic component gives of the idiomatic status of such stretches.

Before showing the manner in which a semantic theory can provide such semantic interpretations for idiomatic stretches, it is necessary to differentiate two sorts of things that are traditionally referred to as idioms. The characterization of an idiom as any concatenation of two or more morphemes whose compound meaning is not compositionally derived from the meanings of the concatenated morphemes does not differentiate those idioms that are syntactically dominated by one of the lowest level syntactic categories, i.e., noun, verb, adjective, etc., from those whose syntactic structure is such that no single lowest level syntactic category dominates them. Let us call the

former type 'lexical idioms', the latter 'phrase idioms'. We shall be concerned primarily with the latter type.

The syntactic component of a linguistic description contains two parts, a set of syntactic rules and a lexicon. The lexicon contains all of the lexical morphemes. The grammatical morphemes are all introduced by the syntactic rules. These rules generate derivations whose final lines contain not lexical items but particular grammatically marked positions in which lexical items may be placed. There is then a substitution condition that permits lexical items from the lexicon to be substituted into derivations provided that the grammatical markings of the lexical item are compatible with those of the grammatical position into which it is substituted. Most of the entries in the lexicon will presumably be single morphemes, e.g., book, run, hate, big, etc., but a signiicant number will be compounds of two or more morphemes, e.g., tele+phone, bari+tone, etc. These are the entries for lexical idioms. They are marked as idioms by virtue of the fact that in the dictionary of the semantic component (not identical with the lexicon of the syntactic component) these sequences of two or more morphemes are directly assigned readings that represent their senses. These readings are not the result of projection rules amalgamating readings for individual morphemes listed separately in the dictionary.

It might be maintained that phrase idioms should also be handled simply by listing each of them as a single lexical item and assigning each a set of readings. This would mean eliminating the distinction between lexical and phrase idioms by also treating the latter in the lexicon as representatives of the lowest level syntactic categories. It would mean that an expression such as kicked the bucket in its idiomatic meaning of 'died' would be regarded as a compound intransitive verb. However, it is easy to show that not all idioms can be regarded as lexical idioms. This follows because there is a large class of idioms, like kicked the bucket, which have a compositional meaning, as well as an idiomatic one. These cannot be regarded as compound lexical items on a par with tele+phone, bari+tone, etc., because listing elements like kicked the bucket in the ordinary lexicon unnecessarily complicates both the syntactic and phonological components of a linguistic description. First, for every case of an idiomatic stretch x which also has a compositional meaning, regarding x as a lexical idiom requires that a new entry be added to the syntactic lexicon (e.g., intransitive verb = kicked the bucket). This addition is quite unnecessary, however, because the syntactic component must already generate x with its atomic parts as lexicon entries in order to provide the formal structure that bears the compositional meaning. Second, the phonological component operates on the syntactic structure of a sentence to assign it a phonetic shape.[5-7] Thus, in English the rules that assign stress patterns to sentences operate on the final derived phrase markers of these sentences. But in the case of idiomatic stretches like kicked the bucket both an occurrence with idiomatic meaning and one with compositional meaning have the

same stress pattern. Therefore all instances of both types of occurrence must have the same syntactic description. Kicked the bucket cannot have in one case the structure Intransitive Verb and in another the structure Verb + Noun Phrase. This latter treatment, if carried out consistently for all such cases, would lead to enormous complications of the phonological component.

Thus elementary considerations of grammatical simplicity suffice to show that at least some phrase idioms cannot be treated as lexical idioms, i.e., as members of one or another of the lowest level syntactic categories. Instead, at least the members of the class of idioms whose occurrences also have compositional meanings must receive the ordinary syntactic structure assigned to occurrences of the stretches with compositional meanings.

The previous considerations establish the fact that the semantic interpretation of sentences containing idiomatic stretches with compositional parallels will have to both account for the idiomatic meaning and mark the semantic ambiguity between this and the compositional meaning. A sentence such as

(1) the old lady kicked the bucket

must be assigned two readings, one of which attributes to it the meaning 'the old lady struck the bucket with her foot', while the other attributes to it the meaning 'the old lady died'. Moreover, the semantic interpretations of sentences containing phrase idioms will have to be properly related to the semantic interpretations of other sentences. Thus sentence (1) must be marked as a paraphrase (on a reading) of the old lady died; the sentence

(2) the man who has been dead for a week just kicked the bucket

must be marked as semantically anomalous on both of its readings, although for different reasons on each, etc.

2. Phrase Idioms

In order for the semantic component to handle phrase idioms properly, it is necessary to broaden the conception of the dictionary subpart of a semantic component presented in previous discussions of semantic theory.[1-4] In addition to having entries for unitary and compound lexical items, the semantic dictionary must also contain entries for the phrase idioms of the language, each such entry associating with its idiomatic stretch a set of readings to represent the sense of that stretch. The dictionary should now be thought of as having two parts, a lexical-item part and a phrase-idiom part. Entries in the latter part will have the form: first, a particular string of morphemes, the idiomatic stretch; next, some associated constituent that must dominate the idiomatic stretch in the phrase markers that are to be assigned the semantic information

associated with the pair of the idiomatic stretch and its dominating constituent, and, finally, this semantic information itself.

Access to the semantic information in the two subparts of the dictionary is obtained by different methods of assigning sets of readings associated with dictionary entries to the minimal semantic elements in the underlying phrase markers of sentences. In the case of minimal semantic elements that are lexical items, i.e., single morphemes or lexical idioms, the method of assigning readings is the following: Given such a minimal element e in an underlying phrase marker M, assign to e in M all and only the readings from e's dictionary entry which are compatible with the syntactic structure of e in M.[8] This method thus associates sets of readings with the lowest level or terminal elements of underlying phrase markers. In the case of minimal semantic elements that are not lexical items, i.e., phrase idioms, the method of assigning readings is the following: Given a string of morphemes t that is dominated by the constituent C in the underlying phrase marker M, assign to the set of readings associated with C in M those readings from the dictionary entry for t that have the form $t \rightarrow C \rightarrow X$, where X is the representation of the meaning of t provided by the dictionary of the semantic theory. This method thus assigns readings to higher level constituents in underlying phrase markers, not to terminal symbols. For example, in the case of the idiomatic stretch kicked the bucket, which has as part of its phrase-idiom dictionary entry the dominating constituent MV, the reading representing 'die' is associated not with any of the morphemes composing this stretch, but rather with the constituent MV that dominates occurrences of kick the bucket in underlying phrase markers. Note that it is this feature of assigning readings representing the meanings of phrase idioms directly to higher level constituents which is the aspect of the semantic theory's treatment of idioms wh. h represents the fact that their meaning cannot be broken up into components and these parcelled out to the morphemes that make up the idiomatic stretch.

Of course, once readings have been supplied for all idioms in a sentence, the projection rules of semantic theory operate in the normal fashion, amalgamating readings drawn from sets of readings associated with constituents to form derived readings to be assigned to the constituent dominating them. A pair of readings, one or both of which is a reading for an idiom, amalgamate in exactly the same manner as readings that are not readings for idiomatic stretches.

Given this approach to the problems of idioms, linguistic theory provides an explication of the difference between sentences whose meaning is wholly a compositional function of the meanings of its lexical items and sentences whose meanings are at least in part determined idiomatically (i.e., by phrase idioms). The meaning of a sentence is idiomatically determined if and only if its semantic interpretation is assigned partly on the basis of information obtained from the phrase-idiom part of the semantic dictionary. The phrase idioms of a language are just those expressions listed in the

phrase-idiom part of the dictionary of the semantic component of the optimum linguistic description of that language.

3. Further Justification for Our Approach

Our approach to phrase idioms is, of course, justified to a great extent by the consid-erations that originally motivated it, i.e., by the fact that for a large class of idioms, namely those with compositional parallels, it eliminates the otherwise required ad hoc additions and unnecessary complications to the syntactic and phonological components. But there are other more subtle justifications. These in turn can also serve to further justify some quite independently motivated features of the syntactic component, and by doing so further justify our approach to phrase idioms.

Consider the following sentences:

(3) John kicked the bucket

(4) the bucket was kicked by John

Clearly sentence (3) is semantically ambiguous. Thus, the semantic component must mark this ambiguity and assign to its terms the readings 'John died' and 'John struck the bucket with his foot'. We have explained the manner in which an extension of the dictionary component, in combination with the ordinary projection rule apparatus, can accomplish this. But the semantic theory must also account for the fact that sentence (4) is definitely unambiguous and is a paraphrase of sentence (3) on the latter reading but not on the former. Sentence (4) is, of course, 'the passive of' sentence (3). In ear-lier transformational descriptions of English it was assumed that sentences like (4) were derived by the action of the passive transformation on the identical phrase marker that underlies sentence (3). Corresponding actives and passives would thus always have identical underlying phrase markers. More recently, however, it has been suggested by Klima[9] that passive sentences have a different underlying phrase marker from active sentences, one that contains a Manner Adverbial constituent represented terminally by a passive morpheme. The passive transformation then substitutes by plus the subject NP of this underlying phrase marker for this passive morpheme, places the object NP where the subject originally was, and adds certain elements to the Auxiliary constituent. This treatment is dictated syntactically by formal considerations within the theory of transformational grammar, chiefly those having to do with the automatic assignment of derived constituent structure by simple, general, mechanical conditions. It also helps to account for certain selectional restrictions on passive constructions, in partic-ular the fact that verbs that do not occur with Manner Adverbials do not have passive forms. The question naturally arises whether or not any external, in particular, any semantic, justification can be found for this way of treating the passive.

Consider again sentences (3) and (4). Taken together, the treatment of the passive

by means of a passive morpheme in underlying phrase markers and the conception of idioms presented above suffice to explain why sentence (3) but not sentence (4) is ambiguous. For note that the entry for kick the bucket in the phrase-idiom component of the semantic dictionary will be of the form: kick+the+bucket → MV → reading that represents the meaning 'die'. That is, in an underlying phrase marker in which the constituent MV dominates kick the bucket, the reading for 'die' is associated with MV. The underlying phrase marker for sentence (3) is
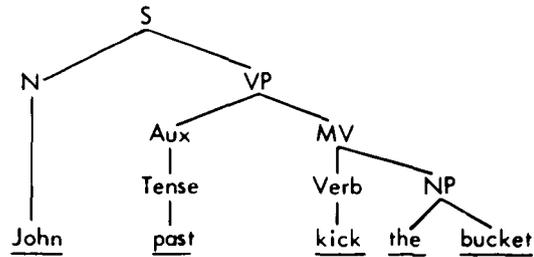
Diagram 1

This phrase marker is such that the conditions for access to semantic information in the phrase-idiom part of the semantic dictionary is met for the stretch kick the bucket, and hence the reading for 'die' is assigned to the constituent MV. But the underlying phrase marker for sentence (4), under the new treatment suggested by Klima, is
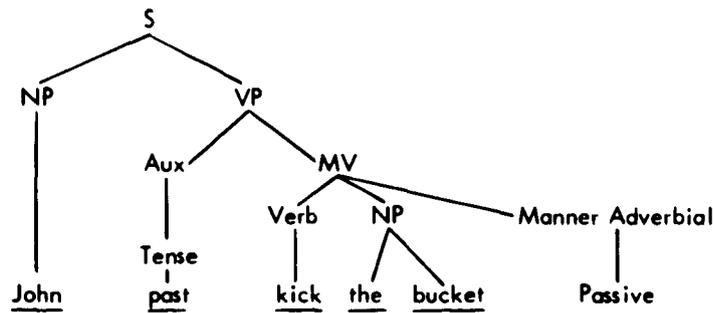
Diagram 2

But here the constituent MV does not dominate the string of morphemes kick+the+bucket as required by the syntactic part of the entry in the dictionary for the phrase idiom kick the bucket, but rather MV dominates kick+the+bucket+Passive. Hence the reading for 'die' cannot be assigned to MV in this phrase marker and thus the sentence that

Diagram 2 underlies, namely (4), is unambiguous although the active that has Diagram 1 as its underlying phrase marker is semantically ambiguous. We thus have the striking result that the treatment of the passive suggested by Klima and the present treatment of phrase idioms automatically explain the fact that the idiomatic meaning of <u>kick the bucket</u> is not found in the passive forms that correspond to this VP. This serves to justify both our treatment of idioms and Klima's syntactic description of the passive.[10]

Note, incidentally, that the lack of ambiguity of sentence (4) cannot be accounted for by claiming that idiomatic meanings are not carried over by transformations. Both sentences (5) and (6) are semantically ambiguous, having one sense that is due to a phrase idiom, and yet one is the result of the question transformation, the other of the imperative transformation:

(5) did John kick the bucket?

(6) kick the bucket!

4. Syntactically Deviant Idioms

Our treatment of idioms leaves open the question of how a linguistic description is to handle idioms that are not syntactically well formed. Because the idiom <u>kick the bucket</u> is syntactically well formed, the syntactic component generates sentences containing occurrences of this idiom and the semantic component can assign them the readings found in their entries in the phrase-idiom part of the dictionary. But idioms that are not syntactically well formed, such as <u>beat about the bush</u>, will not appear as constituents of sentences that are generated by the syntactic component. If these idioms did appear in strings that are generated by the syntactic component, this component would not be empirically adequate because its output would contain some ungrammatical strings. Thus, there will be no occurrences of syntactically deviant idioms available for the normal process of semantic interpretation.

A suggestion of Chomsky offers a way of handling these idioms within the framework of the present paper.[11] He pointed out that sentences containing such idioms can be generated by the device that gives a syntactic description of the semisentences of the language.[12] This being so, if there is in the phrase-idiom part of the dictionary a section containing entries for these syntactically deviant phrase idioms, then these entries can be assigned to occurrences of such idioms in semisentences — so long as some provision is made for the semantic interpretation of semisentences.[13]

J. J. Katz, P. M. Postal

References

1. J. J. Katz and J. A. Fodor, The structure of a semantic theory, Language (in press).

2. J. J. Katz, Analyticity and contradiction in natural languages, Readings in the Philosophy of Language, edited by J. A. Fodor and J. J. Katz (Prentice-Hall, Inc., Englewood Cliffs, N. J., in press).

3. J. J. Katz, The nature of a semantic theory, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M.I.T., April 15, 1963, pp. 181-193.

4. J. J. Katz and P. M. Postal, Towards an Integrated Theory of Linguistics Descriptions, Unpublished manuscript.

5. N. Chomsky, Explanatory models in linguistics, Logic, Methodology and Philosophy of Science: Proc. 1960 International Congress, edited by E. Nagel, P. Suppes, and A. Tarski (Stanford University Press, 1962), pp. 528-550.

6. N. Chomsky, The logical basis of linguistic theory, Proc. Ninth International Congress of Linguists (Mouton and Company, 's-Gravenhage, in press).

7. N. Chomsky and M. Halle, The Sound Pattern of English (Harper and Row, New York, in preparation).

8. The manner in which the notion 'compatible with' is to be explicated in syntactic terms requires an extensive statement of the form of the syntactic component which cannot be given here.

9. E. S. Klima, Unpublished work, 1960-1961.

10. The present treatment of idioms like kick the bucket provides considerable support for the treatment of the tense suffixes of the verb as part of the Auxiliary constituent in underlying P-markers. Clearly simplicity considerations require not separate phrase-idiom entries for kicked the bucket, kicking the bucket, kicks the bucket, etc., but rather only a single entry for kick the bucket. But this is only possible if MV does not dominate the tense morphemes in underlying phrase markers.

11. N. Chomsky, Private Communication, 1963.

12. N. Chomsky, Some methodological remarks on generative grammar, WORD 17, 219-239 (1961).

13. J. J. Katz, Semi-sentences, Readings in the Philosophy of Language, op. cit.

B. STRESS AND PITCH IN THE SERBO-CROATIAN VERB

The stressed syllable in a Serbo-Croatian word can have either rising or falling accentuation. The occurrence of rising and falling accentuation is restricted as follows:

(a) Falling accentuation is possible only on the first syllable of a word; and

(b) Rising accentuation is possible only in words of two or more syllables and can occur on any syllable except the last (thus, in particular, monosyllables can have only falling accentuation).

Suppose that a syllable is called high-pitched if it starts on a high pitch and low-pitched if it starts on a low pitch. Then every word has exactly one high-pitched syllable: if the word has falling accentuation, then the stressed syllable (i.e., the first syllable) is high-pitched, and if it has rising accentuation, then the syllable following the stressed syllable is high-pitched. Thus the location and type (rising or falling) of stress can be predicted from the location of high pitch. Moreover, high pitch can occur on any syllable of the word.[1]

Serbo-Croatian verbs are of two types: thematic and athematic. <u>Thematic</u> verbs are those that have a verbalizing suffix, for example, igrati = igr + aj + ti, where /aj/ is the verbalizing suffix. <u>Athematic</u> verbs are those that have no verbalizing suffix, for example, <u>trēsti</u> = <u>trēs</u> + <u>ti</u>. I assume that the infinitive and present tense forms of thematic verbs have the following immediate constituent structure:

$$\text{infinitive: } \left(\left(\text{stem} + \begin{array}{c}\text{verbalizing}\\\text{suffix}\end{array}\right) + \begin{array}{c}\text{infinitive}\\\text{ending}\end{array}\right)$$

example: igrati = ((igr + aj) + ti)

$$\text{present tense: } \left(\left(\text{stem} + \begin{array}{c}\text{verbalizing}\\\text{suffix}\end{array} + \begin{array}{c}\text{present}\\\text{theme}\end{array}\right) + \begin{array}{c}\text{personal}\\\text{ending}\end{array}\right)$$

example: igrām = ((igr + aj + ā) + m)

There are two types of thematic verbs: those that have the same stress in the infinitive and all forms of the present tense, and those that undergo what Schooneveld[2] calls the "cardinal intonational alternation." Schooneveld defines the cardinal alternation as an alternation between a "rising non-initial penultimate syllable in the infinitive and a rising accent on the preceding syllable in the truncated forms of the present or (if the penultimate syllable is at the same time the initial syllable), between a rising initial-penultimate syllable and a falling initial-penultimate syllable". Examples[3]:

Infinitive	venčãvati	ìgrati
1st sg.	venčãvām	ìgrām

Schooneveld speaks as if there were two separate alternations going on here: one a shift of rising intonation, the other an alternation between rising and falling intonation. However, if high pitch rather than stress is marked in the examples above, it will be noted that in both cases all that happens is that the high pitch is shifted one syllable to the left:

venčãvàti	ìgràti
venčãvàm	ìgrām

(the raised dot denotes high pitch).

The infinitive and present tense forms of a typical verb that undergoes the cardinal alternation are

	phonetic form		underlying form
infinitive	ìgrati	= igràti	((igr + aj) + ti)
1st sg.	igrām	= igrām	((igr + aj + a) + m)
2nd sg.	igrãš	= igrãš	((igr + aj + a) + š)
3rd sg.	igrā	= igrā	((igr + aj + ā) + φ)
1st pl.	igrāmo	= igrāmo	((igr + aj + a) + mo)
2nd pl.	igrāte	= igrāte	((igr + aj + a) + te)
3rd pl.	igrajū	= igràjū	((igr + aj + ā) + ū)

The verbalizing suffix /aj/ is deleted in all present tense forms except the 3rd pl. by rules that will be formulated later; note that in these forms the stem vowel is high-pitched. In the infinitive and the 3rd plural, in which the verbalizing suffix is not deleted, the verbalizing suffix is high-pitched.[4] These facts can be accounted for by saying that the underlying forms of these words have a high-pitched verbalizing suffix and that if the verbalizing suffix is deleted, the high pitch is shifted from it onto the preceding syllable. The appropriate rules for deleting the verbalizing suffix in the five present tense forms in which it is dropped and for deleting the theme vowel $\bar{a}$ in the 3rd person plural seem to be a rule that deletes /j/ between like vowels followed by either a consonant or a word boundary, and a rule that deletes the first of a sequence of two vowels. I thus formulate the rules:

$$3. \quad j \rightarrow \phi \quad \text{in env} \quad \begin{bmatrix} + \text{ voc} \\ - \text{ cns} \\ \alpha \text{ grv} \\ \beta \text{ flt} \\ \gamma \text{ dif} \end{bmatrix} \underline{\qquad} + \begin{bmatrix} + \text{ voc} \\ - \text{ cns} \\ \alpha \text{ grv} \\ \beta \text{ flt} \\ \gamma \text{ dif} \end{bmatrix} \begin{Bmatrix} C \\ \# \end{Bmatrix}$$

$$4. \quad V \rightarrow \dot{V} \quad \text{in env} \quad \underline{\qquad} C_0 \dot{V} V$$

$$5. \quad V \rightarrow \phi \quad \text{in env} \quad \underline{\qquad} V$$

These rules will be assumed to be part of a cycle, i.e., the rules are to be applied first to the innermost immediate constituents, then to the next innermost constituents, etc.

These rules will correctly generate the relevant forms of thematic verbs with cardinal alternation. How then should thematic verbs with fixed stress be handled? For verbs such as v̩ideti/v̩idim, the solution is obvious: the /i/ of the stem is marked high-pitched. Since rule 4 can only affect verbs with a high-pitched verbalizing suffix, the high pitch would remain on the /i/ throughout the entire paradigm.

However, there are also verbs such as čitati/čitām represented in my system as čitati/čitām. If the underlying form for čitati were marked with a high pitch on the /a/, the incorrect form *čitām = *čitām would be obtained, since rule 4 would automatically shift the high pitch onto the first syllable. The solution that I propose is to treat the underlying forms for the paradigm of čitati as having no high-pitched syllables whatever and having a rule that marks the last syllable as high-pitched if none of the preceding syllables are high-pitched:

$$7. \quad V \rightarrow \dot{V} \quad \text{in env} \quad (X \underline{\qquad}), \text{ where X contains no } \dot{V}.[5]$$

For the rule to yield the correct forms čitati and čitām, it will be necessary to have the rule apply as part of the cycle. If the rule were not in the cycle, it of course would give the incorrect form *čitati = *citati. However, if the rule is put in the cycle, it operates as follows:

		infinitive		
		((čit + aj) + ti)	((čit + aj + ā) + m)	((čit + aj + ā) + ū)
first pass through cycle	rules 3-5	no effect	no effect	no effect
	rule 7	čit + aj	čit + aj + ȧ	čit + aj + ȧ
		(čit + aj + ti)	(čit + aj + ȧ + m)	(čit + aj + ȧ + ū)
second pass through cycle	rule 3	no effect	čit + a + ȧ + m	no effect
	rule 4	no effect	no effect	čit + aj + ȧ + ū
	rule 5	no effect	čit + ȧ + m	čit + aj + ū
	rule 7	no effect	no effect	no effect

and thus gives high pitch on the correct syllables.

An alternation between plain and palatalized or palatal consonants, known as transitive softening or jotovanje, takes place in Serbo-Croatian. I am at the moment unprepared to state the exact environment in which it takes place, other than to state the general shape of the rule: the change occurs when a consonant is followed by two vowels that satisfy some condition that I see no way of stating other than to list the combinations of vowels before which the change occurs. It occurs, for example, in the present tense of brisati: ((bris + a + ē) + m) →brišēm. In any event, the simplest way to treat the phenomenon probably will be to say that a /j/ is inserted before the appropriate combination of vowels and that by later rules kj →č, tj →ć, sj →š, etc. Stating the rule in terms of the insertion of a /j/ allows an elegant treatment of the ova/uj alternation. Suppose that /v/ and /j/ are represented in the dictionary forms by /u/ and /i/ (which can be done, since the former segments occur only in postvocalic and initial prevocalic position). Thus the rules will have to contain somewhere a rule that states that high vowels become glides in postvocalic or initial prevocalic position. Suppose that kupovati is represented as ((kup + ou + a) + ti). The rule just mentioned will convert the /u/ into /v/ (actually, into /w/, which later becomes /v/), yielding kupovati. The present tense form kupujem has the expected underlying form ((kup + ou + a + ē) + m). Since a + ē is one of the vowel combinations before which transitive softening takes place, the latter form would be converted into ((kup + ouj + a + ē) + m). If this all precedes rule 5, that rule will delete vowels followed by vowels and yield ((kup + uj + ē) + m), i.e., the correct form.

The transitive softening rule would have to precede the rule for converting /u, i/ into /v, j/, since otherwise it would apply to the /ou/ in the present tense form and yield the incorrect form *kupovjēm → *kupovljēm. I shall assume that these two rules are part of the cycle. Thus rules 3-5 will be preceded by the rules

1. /j/ is inserted in env _____ + V + V, subject to some condition on the V's

2. $\begin{bmatrix} - \text{ cns} \\ + \text{ dif} \end{bmatrix} \rightarrow [- \text{ voc}]$ in env $\left\{ \begin{matrix} V \\ \# \end{matrix} \underline{\hspace{1cm}} V \right\}$

I maintain that the rules given thus far are sufficient to account for the stress shifts in all six classes of thematic verbs given by Meillet and Vaillant.[6]

There are several verbs that Meillet does not include in his six types but that behave like those treated above. Consider, for example, the verbs

| Infinitive | mlẹti | brạti | klạti |
| Ist sg. | meljēm | bẹrēm | kọljēm |

Suppose that base forms mel + ẹ, ber + à, and kol + à are assumed. The 1st sg. will then have the underlying forms ((mel + ẹ + ē) + m), ((ber + à + ē) + m), and ((kol + à + ē) + m). In the first pass through the cycle, transitive softening will occur, the high pitch will be shifted one syllable to the left, and the verbalizing suffix will be eliminated, yielding mèlj + ē + m, berj + ē + m, and kōlj + ē + m. On the second pass through the cycle, none of the rules apply. These are the correct forms for mleti and klati, and in the case of brati, the correct form is obtained by a later rule by which rj → r (/rj/ never occurs in Serbo-Croatian, having coalesced with /r/). The infinitive will have the underlying forms ((mel + è) + ti), ((ber + à) + ti), and ((kol + à) + ti). None of the rules in the cycle apply, so that the forms melèti, kolàti, and beràti, i.e., *mẹleti, *bẹrati, and *kolati, would be obtained. However, if the grammar were provided with a rule that deleted low-pitched /e/ and /o/ in CVL stems, what would remain after the application of that rule is ml + è + ti, br + à + ti, and kl + à + ti, i.e., the correct forms. The incorporation of such a rule into the grammar appears to do no harm, since, as far as I can determine, all C$\left\{ \begin{matrix} e \\ o \end{matrix} \right\}$ L stems work like this.

Consider now verbs with prefixes. In the thematic verbs of the six types that Meillet recognizes, a prefix does not affect the location of high pitch:

stress notation		high-pitch notation	
grādíti	grādīm	grādití	grādīm
razgrādíti	razgrādīm	razgrādití	razgrādīm

The stressed prefix in the present tense of these verbs is thus not a case of stress shift (as it is traditionally described), since the high pitch is on the same syllable regardless of whether or not there is a prefix.

However, the prefix does become high-pitched in the following verb:

| zrẹti | zrēm | (= zrèti | zrēm) |
| obạzreti | obazrēm | (=obazrèti | obàzrēm) |

I will disregard for the moment what happens to the stress and concentrate on

the segmental phonemes. The obvious underlying form for zreti/zrēm is zr + ej (i.e., a nonsyllabic stem). The present tense form zrēm arises from ((zr + ej + ē) + m) by the truncation rules 3 and 5. However, it will be noted that this solution automatically produces the correct stress as well: given the underlying form ((oba + zr + ėj + ē) + m), rule 4 would shift the high pitch from the /e/ of the stem onto the /a/ of the prefix, and the correct form (obȧ + zr + ē + m) would be obtained.

There are a large number of other verbs besides <u>zreti</u> in which the prefix becomes high-pitched in the present tense, for example,

Infinitive	dọbiti	ụmrēti	prosūti	nạdūti	pọpēti
	dọbijēm	ụmrēm	prọspēm	nạdmēm	pọpnēm

A further peculiarity of these verbs is that in the prefixless verbs (or at least, those that exist: many of the stems in question only occur with prefixes) the high pitch in the <u>infinitive</u> is on a different syllable than when there is a prefix (recall that with all verbs treated thus far, even <u>zreti</u>, the high pitch has always been on the same syllable in the infinitive regardless of whether or not there was a prefix). Example: mrēti but ụmrēti (i.e., mrēti but umrėti).

Suppose that the stems <u>mr</u>, <u>sp</u>, <u>dm</u>, <u>pn</u>, etc., are regarded as ending in a high-pitched segment (which in this case will be a consonant) and that the stress shift rule is modified so as only to require a high pitch (not necessarily a high-pitched vowel) in its environment:

4'. V → V̇   in env _____ C_0[+ high pitch] + V.

Then whenever one of the stems <u>mr</u>, <u>sp</u>, <u>dm</u>, <u>pn</u> is followed by a vocalic ending, rule 4' will move the high pitch onto the prefix if there is one. This can also be made to work for <u>biti</u> by representing it with a high-pitched /j/ if rule 4' is modified still further so as to allow an optional vowel between the C_0 and the high-pitched segment. However, the rule still would work correctly after the modification, since verbs of the <u>biti</u> type are the only words in which a vowel would ever be followed immediately by a high-pitched segment.

To generate the infinitives of verbs with consonantal stems, a rule will be needed which inserts a vowel before the infinitive ending and (in the case of stems ending in a true consonant) deletes the stem-final consonant (however, stem-final liquids are retained. To obtain the correct stress, it is necessary that this rule be in the cycle, specifically, between rules 5 and 7, and that rule 7 be modified so that it will not apply to the prefixless infinitives on the first cycle. The obvious modification is to require the constituent to which rule 7 is applying to consist of at least two morphemes, i.e., to say that a vowel bᵥ comes stressed in the environment (X + Y _____), where X and Y contain no V̇. The infinitives mrēti and umrēti are generated as follows (using 6

to denote the rule for modifying nonsyllabic stems):

		((mr̆) + ti)	((u + mr̆) + ti)
first pass through cycle	1-5	no effect	no effect
	6	mr̆ē	u + mr̆ē
	7	no effect	u + mr̆ē̇

		(mr̆ē + ti)	(u + mr̆ē̇ + ti)
second pass through cycle	1-6	no effect	no effect
	7	mr̆ē + ti	no effect

Final result: mr̆ēti, umr̆ēti.

Athematic verbs can be either nonsyllabic or monosyllabic. The rules formulated thus far generate correctly the relevant forms of all nonsyllabic athematic verbs. Let me now turn to the monosyllabic ones. Monosyllabic athematic verbs fall into three classes:

1) those that have rising intonation in the infinitive and all present tense forms: trēsti/trēsēm.

2) those that have a short falling pitch in the infinitive but a long rising pitch in the present tense: ᷉risti/grīzēm, and

3) three verbs leći, reći, and moći, which have a rising pitch in the infinitive but a falling pitch in the present tense.

I see no alternative to simply treating type (3) as exceptions.

The obvious solution for type (1) is to represent the stems as having no high pitch. The rule 7 would put the high pitch on the -ti of the infinitive and the theme vowel ē of the present tense. This gives the correct answer for all forms except one, namely the 3rd person plural. The expected underlying form ((trēs + ē) + ū) for the third person plural trēsū of trēsti would come out of the first pass through the cycle with a high pitch on the ē. On the second pass, rule 4 would shift the high pitch to the preceding syllable, rule 5 would delete the ē, and the incorrect form *trēsū would be obtained. The only solutions that I can think of to this difficulty are all somewhat ad hoc, and all are essentially equivalent to saying that the present tense does not have the IC structure ((stem + theme) + person) but rather (stem + theme + person). If such a structure is assumed, then the relevant forms will go only once through the cycle, and rule 7 will put the high pitch on the final vowel; in particular, in the 3rd person plural, it will put the high pitch on the ū. However, I am unable to find any independent justification for assuming that athematic verbs have a different IC structure from thematic verbs.

Verbs of group (2) will undoubtedly require some special rule, since I know of no

process of vowel shortening or lengthening or stress shift of which they could be considered a special case (note that in verbs of group (2) the stress shift is in the opposite direction from that in cardinally alternating thematic verbs). The simplest ad hoc rule that I can think of for this alternation is a rule that makes a high-pitched vowel long and low-pitched in a monosyllabic stem followed by a vowel. That would involve representing type (2) stems as having a short high-pitched vowel, which is possible, since there are no other monosyllabic athematic verbs which there would be any reason to represent with a high-pitched vowel.

The order of application of the rules arrived at above is the following:

1. Insert /j/ in env _____ + V + V, subject to some condition on the V's.

2. $\begin{bmatrix} - \text{ cns} \\ + \text{ dif} \end{bmatrix} \rightarrow [- \text{ vos}]$ in env $\left\{ \begin{matrix} V \\ \# \end{matrix} \underline{\qquad} V \right\}$

3. $/j/ \rightarrow \phi$ in env $\begin{bmatrix} + \text{ voc} \\ - \text{ cns} \\ \alpha \text{ grv} \\ \beta \text{ flt} \\ \gamma \text{ dif} \end{bmatrix} \underline{\qquad} \begin{bmatrix} + \text{ voc} \\ - \text{ cns} \\ \alpha \text{ grv} \\ \beta \text{ flt} \\ \gamma \text{ dif} \end{bmatrix} \left\{ \begin{matrix} C \\ \# \end{matrix} \right\}$

4. $V \rightarrow \dot{V}$ in env _____ $C_0(V)$ [+ high pitch] + V

5. $V \rightarrow \phi$ in env _____ V

6. Modification of nonsyllabic stems before the infinitive ending (an $\bar{u}$ or $\bar{e}$ is added after the stem; if the stem ends in a true consonant, the latter is deleted).

7. $V \rightarrow \dot{V}$ in env (X + Y_____), where X and Y contain no $\dot{V}$.

<div align="right">J. D. McCawley</div>

References

1. The possibility of representing stress in Serbr-Croatian by marking the occurrence of high pitch was discovered independently by myself and E. Wayles Browne of Harvard University

2. C. H. van Schooneveld, Serbo-Croatian conjugation, International J. Slavic Linguistics and Poetics 1, 55-69 (1957).

3. Instead of the traditional notation, I will use the more suggestive notation of marking rising and falling intonation by ∕ and ∖ , respectively, beneath the vowel. My notation and the traditional notation are related as follows:

	my notation	traditional notation
short rising	a̗	à
long rising	a̱̗	á
short falling	a̖	à
long falling	a̱̖	â

The notation used here is a variant of that used in H. G. Lunt, On the study of Slavic accentuation, WORD 19, 82-99 (1963).

4. It is interesting to note that in Russian the basic stress moving rule calls for stress to be shifted when a stressed vowel is deleted, much like this rule. However, in Russian the direction of shift is the opposite, namely to the <u>right</u> rather than to the left.

5. The parentheses denote the beginning and end of the constituent to which the rule is applying.

6. A. Meillet and A. Vaillant, <u>Grammaire de la Langue Serbo-Croate</u> (Champion, Paris, 2d edition, 1952).

C. SOME LANGUAGES THAT ARE NOT CONTEXT-FREE

In many frameworks for describing the grammar of natural languages it has been assumed, implicitly or explicitly, that the grammar will assume the form of what Chomsky[1-3] has termed "context-free grammar." We shall consider a subpart of English (and many other natural languages) which is demonstrably <u>not</u> context-free.

Consider the English cardinal numbers. There are an infinite number of them, and each is of the form

$$N \; T^n \; (, \; N \; T^{n-1}) \; \ldots \; (, \; N \; T) \; (, \; N) \tag{1}$$

where the comma indicates a comma intonation, the parentheses indicate that the enclosed elements are optional, T^n abbreviates n occurrences of <u>thousand</u>, and N indicates a number between 1 and 999. Here, <u>million</u> is interpreted as <u>thousand thousand</u>, <u>billion</u> as <u>thousand thousand thousand</u>, etc. A closely related analysis of the cardinal numbers treats each of them in a full form

$$N \; T^n, \; N \; T^{n-1} \; \ldots, \; N \; T, \; N \tag{2}$$

where N now indicates a number between 0 and 999. The expression of cardinal numbers in terms of powers of 10 is also of this form, with the comma indicating a comma intonation followed by <u>plus</u>, T abbreviating <u>times ten</u>, and N indicating a number between 0 and 9.

The set of cardinal numbers is of particular interest in a discussion of the form of grammar, since this set is the clearest example of an infinite subpart of natural languages. In the theory of transformational grammars, for example, there is some question as to whether the phrase-structure component of the grammar should generate a finite or infinite set of strings. It is interesting to note that there are quite natural ways of generating the cardinal numbers by means of transformations — in fact, there is a fairly simple set of transformations that generate the cardinal numbers <u>in order</u>.

To investigate whether the set of all English cardinals is CF (context-free) or not, we consider the following sets of strings:

$$P = \left\{ x \; \middle| \; x = b^n (ab^{n-1}) \; \ldots \; (ab^2) \; (ab), \quad \text{where } n = 1, \, 2, \, 3, \, \ldots \right\}$$

$$C = \left\{ x \mid x = b^n ab^{n-1}a \ldots ab^2 ab, \quad \text{where } n = 1, 2, 3, \ldots \right\}$$

It is clear that P is CF if and only if the set of all strings of form (1) is CF, and that C is CF if and only if the set of all strings of form (2) is CF. Along with P, we consider the sets P_i

$$P_i = \left\{ x \mid x = b^{n_i} ab^{n_i-1}a \ldots ab^{n_o}, \quad \text{where } 1 \leqslant n_o < n_1 < \ldots < n_i \right.$$
$$\left. \text{and } n_o = 1, 2, 3, \ldots \right\}, \quad i = 0, 1, 2, \ldots$$

Note that $P = \bigcup\limits_{i \geqslant 0} P_i$. Along with C, we consider the sets C_i.

$$C_i = \left\{ x \mid x = b^{n+i} ab^{n+i-1}a \ldots ab^n, \quad \text{where } n = 1, 2, 3, \ldots \right\},$$
$$i = 0, 1, 2, \ldots$$

It is easy to show that $C_o (= P_o)$, C_1, and P_1 are CF.

We assume that any CF grammar G is such that each symbol in the vocabulary of G occurs in some S-derivation and is such that any S-derivation can be terminated by a finite number of applications of the rules of G. It can be shown that these two well-formedness conditions have no effect upon the set of all CF languages. Moreover, it is easy to show that any infinite CF language has a subset of the form F:

$$zx^m ty^m w \qquad \text{for } m = 1, 2, 3, \ldots$$

where z, x, t, y, and w are fixed strings, t is non-null, and either x or y is non-null.

THEOREM 1: C_i (i$\geqslant$2) and C are not CF.

PROOF 1: Suppose that C_i is CF. Then C_i has a subset of form F. That is,

$$zx^m ty^m w = b^{n_m+i} a \ldots ab^{n_m}$$

so that $z = b^u$ and $w = b^v$, where u and v are fixed and greater than or equal to zero, and so that there must be an a in either x, t or y. If there is an a in x or y, then there are at least m a's in $zx^m ty^m w$. But there are only i a's in $b^{n_m+i} a \ldots ab^{n_m}$. By choosing m = i, we obtain a contradiction.

Hence, all of the a's must occur in t; that is,

$$t = ab^{n_m+i-1} a \ldots ab^{n_m+i} a \qquad \text{for } m = 1, 2, 3, \ldots$$

For i $\geqslant$ 2, as m increases without limit, n_m and hence the length of t also increase without limit. But t must be fixed. From this contradiction, we see that C_i (i$\geqslant$2) is not CF.

The proof that C is not CF is similar; again, x and y must be free of a, so that

$$t = ab^{n_m - 1} a \ldots ab^2 a \qquad \text{for } m = 1, 2, 3, \ldots$$

which is impossible, since t must be fixed.

The method of Proof 1 cannot succeed for P_i and P because P_i and P <u>do</u> have infinite subsets of form F; for example, both P_0 and P_1 are subsets of P, and

$$\{x \mid x = b^m ab^2 ab, \quad \text{where } m = 3, 4, 5, \ldots\}$$

is a subset of P_2. In fact, each P_i ($i \geq 1$) can be expressed as the union of an infinite number of disjoint infinite sets, each of which is of form F. Then, since P is the union of the P_i's, and the P_i's are disjoint, P can be similarly expressed.

In the following proof we assume that for each nonterminal symbol A in a CF grammar G, there are infinitely many strings (hence, infinitely many terminal strings) that are derivable from A. Otherwise, whenever there is a rule $B \rightarrow \phi_1 A \phi_2$ and only $z_1, \ldots, z_n$ are derivable from A, then A can be eliminated from the grammar by replacing the rule $B \rightarrow \phi_1 A \phi_2$ by the rules $B \rightarrow \phi_1 z_i \phi_2$ for $1 \leq i \leq n$.

Suppose that there is a CF grammar G for which $L(G)$ is either P or P_i ($i \geq 2$). We make the following definitions: V_1 is the set containing b and every nonterminal symbol A for which every terminal derivative (i.e., every derivative consisting entirely of terminal symbols) of A consists entirely of b's; T is the set of all strings of symbols in V_1; V_2 is the set consisting of every nonterminal symbol having at least one terminal derivative containing at least one a; V_3 ($\subseteq V_2$) is the set consisting of every nonterminal symbol having an infinite number of terminal derivatives containing two or more a's.

(a) In any S-derivative in G, no nonterminal symbol appears to the right of the symbol a or to the right of a symbol in V_2.

If there were such a nonterminal symbol, then there would be terminal S-derivatives not in P or P_i. Then

(b) If $A \rightarrow \phi$ in G, then ϕ is of the form αBx, where α is null or $\alpha \in T$, B is null or $B \in V_2$, and x is null or terminal.

(c) There is a rule of the form $S \rightarrow \alpha_1 B_1 x_1$ in G, where α_1 and x_1 have the same conditions as α and x, respectively, in (b), and $B_1 \in V_3$.

PROOF OF (c): Consider all rules $S \rightarrow \phi$ in G. By (b), ϕ is of the form αBx. Suppose that in every such rule, B is either null or has only a finite number of terminal derivatives containing two or more a's. Therefore, the terminal derivatives from each αBx are all of the form $b^k y$ or $b^{k_1} ab^{k_2} y$ (for some $k, k_1, k_2 \geq 0$), where y is one of a <u>finite</u> number of fixed terminal strings. But then $L(G)$ is properly included in P or P_i ($i \geq 2$). Then (c) follows from this contradiction.

Consider all of the rules $B_1 \rightarrow \phi$ in G. Again by (b), ϕ is of the form αBx. Suppose that in each such rule B is null or has only a finite number of terminal derivatives

containing two or more a's. We can construct a CF grammar G_1 in the following way: Replace the rules $B_1 \rightarrow aBx$ in G by the rules $S \rightarrow a_1 aBxx_1$ and the rules $\overline{B}_1 \rightarrow aBx$, where $\overline{B}_1$ is a new symbol, and replace any occurrence of B_1 in the remaining rules by $\overline{B}_1$. Then $L(G_1) = L(G)$, but there is no rule satisfying the conditions in (c). From the contradiction, there must be a rule of the form $B_1 \rightarrow a_2 B_2 x_2$ in G, where $B_2 \in V_3$.

We can treat B_2 as we have treated B_1, so as to obtain a $B_3 \in V_3$, etc. There are, however, only a finite number of vocabulary symbols in G, so that either $S \Rightarrow \beta Sy \Rightarrow \beta^n Sy^n$ for some β (null or in T), for some y (null or terminal), and for all $n \geq 2$; or else $S \Rightarrow \beta B_m y \Rightarrow \beta y B_m zy \Rightarrow \beta y^n B_m z^n y$, with appropriate conditions on β, y, z, y, and n. Then x cannot be null, for if it were $L(G)$ would be properly included in P or P_i again. Nor can x = a, since there is no string in P or P_i with two adjacent a's. Hence x contains at least one b. Now S and B_m must both be in V_3, so that they have terminal derivatives containing at least two a's. But then there must be strings in $L(G)$ which are not in P or P_i (i≥2). That is, any CF grammar either is insufficient to generate P or P_i, or else generates strings not in P or P_i. Therefore

THEOREM 2: P and P_i (i≥2) are not CF.

A. M. Zwicky, Jr.

References

1. N. Chomsky, Context-free grammars and pushdown storage, Quarterly Progress Report No. 65, Research Laboratory of Electronics, M.I.T., April 15, 1962, pp. 187-194.

2. N. Chomsky, Formal properties of grammars, Handbook of Mathematical Psychology, Vol. 2, edited by R. R. Bush, E. H. Galanter, and R. D. Luce (John Wiley and Sons, Inc., New York, in press).

3. N. Chomsky, On certain formal properties of grammars, Information and Control 2, 137-167 (1959).

D. NASAL DIPHTHONGS IN RUSSIAN

We define a <u>nasal</u> <u>diphthong</u> as any $\begin{bmatrix} +voc \\ -cns \end{bmatrix} \begin{bmatrix} +cns \\ +nsl \end{bmatrix}$ cluster in the environment _____ $+\begin{Bmatrix} C \\ \# \end{Bmatrix}$. The existence of nasal diphthongs in Russian is

shown in such forms as the following:

Phonetic	Phonemic[1]	English
žát,	$(gím+tí)^2$	to press, squeeze
žm,ǯt	$((glm+é)+tu)$	he presses, squeezes
žát,	$(gín+tí)^2$	to reap
žn,ǯt	$((gln+é)+tu)$	he reaps
nač,æt,	$(na+=+kín+tí)^2$	to begin
nač,n,ǯt	$((na+=+kln+é)+tu)$	he begins
zná̇m,ə	$(znóm+en+\emptyset)$	banner (nom. sg.)
zná̇m,ɪn,ɪ	$(znóm+en+í)$	banner (gen. sg.)
znam,ǯnə	$(znōm+én+ō)$	banner (nom. pl.)

Forms such as those listed above are sufficient to prove the existence of morpho-phonemic front nasal diphthongs in Russian.[3]

It is commonly considered that in contemporary Russian there are no back nasal diphthongs (except, perhaps, for the often quoted but entirely inconclusive forms [zvúk] 'sound' and [zvón] 'peal, ringing, chime'). In the Third Person Plural of verbs, however, we find a perfectly clear case for the existence not only of front nasal diphthongs but of back nasal diphthongs as well.

If we specify the Third Plural morpheme as

$$3 + Pl \quad \rightarrow \quad \underline{n} + \underline{tu}$$

then the phonemic transcription of some representative Third Plural verb forms will be as follows:

Phonetic	Phonemic	English
s,ɪd,á́t	$((síd+é+í)+n+tu)$	they sit
gəvar,á́t	$((gouor+í+í)+n+tu)$	they talk
n,ɪsút	$((nes+é)+n+tu)$	they carry
zná́jut	$((znói+e)+n+tu)$	they know
žmút	$((glm+é)+n+tu)$	they press, squeeze

In order to derive the phonetic transcriptions from the phonemic representations

we require the following rules (note that the rules given below represent simply an extension and a slight reordering of the rules given in [RMCR]; we have ommited all [RMCR] rules that play no part in the present report):

C-1 $\begin{bmatrix} -cns \\ +diff \\ -tns \end{bmatrix}$ → [-vocalic] in env: _____ V

C-2 V → φ in env: _____ + V

C-3 Erase parentheses and return to C-1. If there are no more parentheses, then proceed to P-1.

P-1 $\begin{bmatrix} +obst \\ +cmp \end{bmatrix}$ → $\begin{bmatrix} +str \\ -grv \end{bmatrix}$ in env: _____$\begin{bmatrix} -cns \\ -grv \end{bmatrix}$

P-2 $\underline{e}$ → $\underline{o}$ in env: _____ [+nasal][+cons]^4

P-3 [+cons] → [+sharp] in env: _____$\begin{bmatrix} -cns \\ -grv \end{bmatrix}$

P-4 $\begin{Bmatrix} [+grave] & → & \overline{\underline{u}} \\ \\ [-grave] & → & \underline{\overline{o}} \end{Bmatrix}$ in env: $\begin{bmatrix} \underline{\quad} \\ +voc \\ -cns \end{bmatrix}$[+nasal][+cons]

P-5 $\begin{bmatrix} -voc \\ -obs \end{bmatrix}$ → φ in env: _____ + [+cons]

P-6 $\begin{bmatrix} +voc \\ -cns \\ +diff \\ -tns \end{bmatrix}$ → φ^5

P-7 $\underline{w}$ → $\underline{v}$

We apply these rules to some of the phonemic representations given above:

sidjat: ((sĪd+ē̆+Ī́)+n+tu) →C-2→ ((sĪd+Ī́)+n+tu) →C-3→

    (sĪd+Ī́+n+tu) →C-3→ sĪd+Ī́+n+tu →P-3→ s,Īd,+Ī́+n+tu

    →P-4→ s,Īd,+ó̆+n+tu →P-5→ s,Īd,+ó̆+tu →P-6→ s,Īd,

    +ó̆+t → s,ɪd,át

govorjat: ((gouor+Ī+Ī́)+n+tu) →C-1→ ((gowor+Ī+Ī́)+n+tu)

    →C-2→ ((gowor+Ī́)+n+tu) →C-3→ (gowor+Ī́+n+tu) →C-3→

    gowor+Ī́+n+tu →P-3→ gowor,+Ī́+n+tu →P-4→ gowor,+ó̆+n+tu

    →P-5→ gowor,+ó̆+tu →P-6→ gowor,+ó̆+t →P-7→ govor,+ó̆+t

    → gəvar,át

nesut: ((nes+ɛ́)+n+tu) →C-3→ (nes+ɛ́+n+tu) →C-3→ nes+ɛ́

    +n+tu →P-2→ nes+ó́+n+tu →P-3→ n,es+ó́+n+tu →P-4→

    n,es+ú̆+n+tu →P-5→ n,es+ú̆+tu →P-6→ n,es+ú̆+t →

    n,ɪsút

znajut: ((znó́i+e)+n+tu) →C-1→ ((znó́j+e)+n+tu) →C-3→

    (znó́j+e+n+tu) →C-3→ znó́j+e+n+tu →P-2→ znó́j+o+n+tu

    →P-4→ znó́j+ū+n+tu →P-5→ znó́j+ū+tu →P-6→ znó́j+ū+t

    → znɛ́jut

načala: (na+=+kln+l+ó́) →C-3→ na+=+kln+l+ó́ →P-1→ na+=

    +čln+l+ó́ →P-3→ na+=+č,ln+l+ó́ →P-4→ na+=+č,ōn+l+ó́

    →P-5→ na+=+č,ō+l+ó́ → nač,ɪlá (Fem. Past)

načnut: ((na+=+kln+ɛ́)+n+tu) →C-3→ (na+=+kln+ɛ́+n+tu) →C-3→

    na+=+kln+ɛ́+n+tu →P-1→ na+=+čln+ɛ́+n+tu →P-2→ na+=

    +čln+ó́+n+tu →P-3→ na+=+č,ln+ó́+n+tu →P-4→ na+=

    +č,ln+ú̆+n+tu →P-5→ na+=+č,ln+ú̆+tu →P-6→ na+=+č,n+ú̆+t

    → nač,nút

<div align="right">T. M. Lightner</div>

Footnotes

1. For an explanation of the abbreviations used in our phonemic transcription, see T. M. Lightner, Remarks on the morphophonemic component of Russian (henceforth [RMCR]), Quarterly Progress Report No. 69, Research Laboratory of Electronics, M.I.T., April 15, 1963, pp. 193-199.

2. We require the presence of /i/ in the roots /kin/, /gim/, /gin/ not only to account for the shift of k̲ and g̲ to c̲, and z̲, but also to account for the tense /i/ in derived imprefectives such as nac̲inat̲' 'to begin,' otz̲imat̲' 'to wring out,' doz̲inat̲' 'to finish reaping,' and for the nondiffuse grave /o/ in derived nominals such as konec 'end.'

3. For further examples see T. M. Lightner, On obrazovat̲, and pon, at, type verbs in Russian, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M.I.T., October 15, 1962, pp. 177-180.

4. The rule stated here is actually more general because e̲ → o̲ also before l̲ followed by consonant. Thus, e.g., we shall derive z̲olob from /gelb/, and we shall account for the e̲/o̲ alternation in Inf. mo̲lot' but 3 Sing. me̲let by deriving both forms from the verb stem /me̲l/: Inf. (mel + ti), 3 Sing. ((mel + e) + tu). See Roman Jakobson, Remarques sur l'évolution phonologique du russe comparée à celle des autres langues slaves, III, Sec. 3, p. 21, Travaux du Cercle Linguistique de Prague, II (1929).

5. This rule, of course, will be preceded by a rule that lowers {u̲, i̲} in "strong" position. These two rules will account for such alternations as son (nom. sg.)/sna (gen. sg.), krasen (masc. short)/krasna (fem. short), etc. See Section XVIII-E.

E. NOTES ON THE VERBS čitat' AND -čest'

Both verbs čitat' and -čest' are formed from the root /kit/. The Imperfective requires the verb suffix /ōj/ and root vowel length: /kīt + ōj/. The Perfective requires no verb suffix and retention of the short vowel. The forms are thus as follows:

	Imperfective		**Perfective**	
Inf:	č,ıtát,	(kīt+ój+tī)	č,és,t,	(kít+tī)
Masc. Past:	č,ıtál	(kīt+ój+l+os)	č,ól	(kít+l+os)
Fem. Past:	č,ıtálə	(kıt+ój+l+ō)	č,lá	(kıt+l+ó)
3 Sg:	č,ıtájıt	((kīt+ój+e)+tu)	č,t,ót	((kıt+é)+tu)
3 Pl:	č,ıtájut	((kīt+ój+e)+n+tu)	č,tút	((kıt+é)+n+tu)
Imper:	č,ıtáj	((kīt+ój+e)+ī+#)	č,t,í	((kıt+é)+ī+#)

In addition to the rules already formulated in previous reports,[1] we shall require two more rules. One rule will lower "strong" {u̲, i̲}; the other rule will eliminate "weak" {u̲, i̲}. These two rules will account for both the presence of root vowel in the Masc. Past č,ól and the absence of root vowel in the Fem. Past č,lá. The two rules are as follows:

$$\text{A:} \quad \begin{Bmatrix} \underline{u} \\ \underline{i} \end{Bmatrix} \rightarrow \begin{Bmatrix} \underline{o} \\ \underline{e} \end{Bmatrix} \quad \text{in env:} \underline{\qquad} C_0 \begin{Bmatrix} \underline{u} \\ \underline{i} \end{Bmatrix}$$

B: $\left\{\begin{array}{c} \underline{u} \\ \underline{i} \end{array}\right\}$ → φ

Thus the Masc. and Fem. Past will have the following derivations:

Masc: (kĭt+l+os) → kĭt+l+os → čĭt+l+os → č,ĭt+l+os →

č,ĭt+l+us → č,ĭt+l+u →A→ č,ét+l+u →B→ č,ét+l →

č,ót+l → č,ó+l → č,ól

Fem: (kit+l+ó) → kit+l+ó → čit+l+ó → č,it+l+ó →B→

č,t+l+ó → č,+l+ó → č,lá

It can be seen, however, that no modification of rule A will predict the retention of root vowel in the Infinitive. I believe, therefore, that in addition to rule A one must also posit

A': $\left\{\begin{array}{c} \underline{\acute{u}}'' \\ \prime \\ \underline{i} \end{array}\right\}$ → $\left\{\begin{array}{c} \acute{o} \\ \prime \\ \underline{e} \end{array}\right\}$

The Infinitive will now be derived in the following manner:

Inf: (kĭt+tĭ) → kĭt+tĭ → čĭt+tĭ → č,ĭt+t,ĭ →A'→ č,ét+t,ĭ

→ č,ét+t, → č,és+t, → č,és,+t, → č,és,t,

There is no difficulty in any of the Imperfective forms except for the rule that lengthens the root vowel of /kit/ to /kīt/. Although at present I can give no precise formulation of this rule, I think that the lengthening of this vowel must be accounted for by the same rule that lengthens {u̱,i̱} in derived Imperfectives like <u>nazyvat'</u> 'to call' (from /zu̱v/, cf. Perfective <u>nazvat'</u>) and <u>dobirat'</u> 'to gather' (from /bi̱r/, cf. Perfective <u>dobrat'</u>).

<div align="right">T. M. Lightner</div>

Footnotes

1. See Section XVIII-D, and T. M. Lightner, Remarks on the morphophonemic component of Russian, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M.I.T., April 15, 1963, pp. 193-199.

F. THE SHIFT OF s TO x IN OLD CHURCH SLAVONIC VERB FORMS[1]

It is well known that Indo-European s preceded by i̱, u̱, r̲, k shifted to x̲ in Proto-Slavic.[2] In this report we mention a few OCS verb forms in x̲ (ṣ before front vowels)

for which a synchronic analysis of OCS must posit phonemic s̲ and thus reflect the historic s̲ →x̲ sound shift.

1. **The 2 Sing. Pres. Tense ending:**

Phonetic	Phonemic	English
dasi	(dōd+sĭ)	you give
moliši	((mol+ĭ+ĭ)+sĭ)	you beg
mьniši	((min+ē+ĭ)+sĭ)	you think
glagolješi	((golgol+ō+e)+sĭ)	you speak
miluješi	((mĭl+ou+ō+e)+sĭ)	you pity
etc.		

2. **Aorist:**

1 Sg:	molixъ	((mol+ĭ+s)+u)
2 Pl:	moliste	((mol+ĭ+s)+te)
3 Pl:	moliše	((mol+ĭ+s)+in)
1 Sg:	glagolaxъ	((golgol+ō+s)+u)
2 Pl:	glagolaste	((golgol+ō+s)+tə)
3 Pl:	glagolaše	((golgol+ō+s)+in)
etc.		

3. **Imperfect:**

1 Pl:	moljaaxomъ	((mol+ĭ+ēō+s)+e+mu)
2 Pl:	moljaašete	((mol+ĭ+ēō+s)+e+te)
3 Pl:	moljaaxǫ	((mol+ĭ+ēō+s)+e+on)
etc.		

In the Imperfect there are no telling forms with phonetic s̲ (cf., however, 2 Sing. Pres. da̲s̲i and 2 Pl. Aorist moli̲s̲te), but the analysis with s̲ rather than x̲ must be preferred because then the s̲ will be common to both Past Tense forms. We find external confirmation for the correctness of this solution in historically younger forms that have -sa̲s̲ta/-sa̲s̲te/-sa̲s̲te in the 2 Dual/3 Dual/2 Plural Imperfect.[3]

The rule for the shift of s̲ to x̲ is as follows:

$$\underline{s} \;\rightarrow\; \underline{x} \qquad \text{in env:} \qquad [\text{-obstruent}] + \underline{\qquad} [\text{-obstruent}]$$

This rule explains the retention of $\underline{s}$ in the 2 Sing. Pres. of athematic verbs ($/d\bar{o}\underline{d}+s\bar{\imath}/$) but not in the 2 Sing. Pres. of thematic verbs ($/\ldots+\underline{i}+s\bar{\imath}/$ or $/\ldots+\underline{e}+s\bar{\imath}/$), and the retention of $\underline{s}$ in the 2 Dual, 3 Dual, 2 Plural Aorist ($/\ldots+s+\underline{ta}/$ or $/\ldots+s+\underline{te}/$) but not in the other Aorist forms ($/\ldots+s+\underline{u}/$, $/\ldots+s+\underline{in}/$, etc.). Furthermore, this rule explains the otherwise inexplicable "reappearance" of $\underline{s}$ in the younger -æasta, -æaste (opposed to the older -æašeta, -æašete).

We shall not dwell on the already well-known fact that the environment of this rule must be expanded to include pre-$\underline{s}$-velars because of (to mention but one example) Aorist forms like ræxъ, ræste, ræše from the root /rek/.

<div align="right">T. M. Lightner</div>

Footnotes

1. This report is extracted from a larger work on OCS morphophonemics presented at Linguistics Seminars, Research Laboratory of Electronics, M.I.T., March 13, 1962 and July 15, 1962.

2. See, e.g., Roman Jakobson, Comparative Slavic phonology, Selected Writings (Mouton and Company, 's-Gravenhage, 1962), p. 413.

3. See, e.g., H. G. Lunt, Old Church Slavonic Grammar (Mouton and Company, 's-Gravenhage, 1959), sec. 9.12, p. 86. According to G. Nandriş, Old Church Slavonic Grammar (Athlone Press, London, 1959), p. 147, Sava's Book (Savvina kniga) and the Ostromirovo Evangelie use the -æasta/-æaste forms exclusively, whereas the Codex Zographensis and the Glagolita Clozianus use only the -æšeta/-æašete forms.

G. ALTERNATION OF RULES IN CHILDREN'S GRAMMAR

1. Problem and Method

The process by which language is acquired has been postulated to be stimulus-response-reinforcement associations[1] or drive-cue response-reinforcement associations.[2] Presumably, the environment creates the drive or need to imitate a model. This drive at some stage in development is cued by particular utterances. These utterances are imitated and the response is reinforced either by reiteration, some kind of response from the model or internal gratification. The data (language production or preception) are analyzed according to the particular theoretical point of view held. They are also described simply in terms of traditional linguistic labels. Language is divided into tact and mands.[1] Language is analyzed in terms of cue-response-reward situations.[2] From the purely descriptive point of view the acquistion of phonemes by the infant and young child[3] and the usage of differently structured sentences at various age levels have been described.[4]

Essentially the same conclusions are reached from research undertaken from the

theoretical viewpoints mentioned above and studies using labelling procedures. As the child matures the proportion of usage of more complex language increases. He proceeds from the simple to the complex. At the level of meaningful utterances, the proporation of usage of more complex sentences increases, and, therefore, sentence length increases. Complexity is usually intuitively defined as more difficult and, in a circular fashion, seems, in large part, to be dependent on an expanding lexicon and increasing sentence length.

This study was undertaken to attempt to formalize the notion of increasing complexity in grammar as children mature and in this way to examine further the hypothesis of imitation in language acquistion. The model of grammar used in this study is a generative or transformational model. It is hypothesized that the perceiver or child has incorporated both the generative rules of the grammar and a heuristic component that samples an input sentence and by a series of successive approximations determines which rules were used to generate this sentence.[5-7] Instead of memorizing every sentence that he has been exposed to and imitating these sentences, he uses a set of rules to generate not only the sentences that he has heard, but also other possible examples.

The population in this study was comprised of 159 children, ranging in age from 2 years, 10 months to 7 years, 1 month. They were homogeneous in socio-economic status and I.Q. Language was elicited and recorded in various stimulus situations: 1) responses to a projective test, 2) conversation with an adult, and 3) conversation with peers. The last two situations took place both in controlled and free, that is, classroom, environments. The language sample produced by each child was analyzed by using the transformational model previously described. A grammar was written which included all of the postulated rules used to generate the sentences in the total language sample.[8]

It was found that all of the basic structures used by adults to generate their sentences which we have thus far been able to describe are in the grammar of children from 3 to 4 years of age. Three developmental trends were observed. The first was the very rapid acquisition of some phrase-structure rules that were excluded by some children under three. The second was that increasing numbers of the children used certain transformations as an increasingly older population was observed. The third was the use of alternate rules at all three levels of the grammar.[9] That is, in terms of this model of a grammar, the children generated their sentences from rules that produced completely grammatical structures and simultaneously from rules that did not. These latter rules have been termed rules restricted to a children's grammar. This report deals with this last developmental trend.

Three facts should be kept in mind about the use of these restricted forms. Some of these rules can be found in the grammar of some adults, although it seems unlikely that they are in the grammar of the parents of these children, since all parents have

occupations that fall within the upper 24 per cent of a middle class population. The sentences produced from these rules were deviations from complete grammaticalness but were not completely outside a set of possible sentences. For example, the child might generate the sentence 'You can't put no more water in it.' from these rules but not the nonsentence 'More water no put can't you in it.' or any such departures. Finally, these restricted forms occur infrequently in the total language sample. However, I believe that a close examination of the use of these rules does bring clarification and greater formality to the notion of children using an increasingly complex grammar as they mature.

2. Results

As was stated before, the use of alternate rules occurs at all three levels of the grammar. At the phrase-structure level almost all of the children use verbs, nouns, articles, prepositions, and particles correctly. At the same time, some of the children omit these parts of speech or use them redundantly. Also, substitution of verbs, articles, and prepositions takes place. As examples, verbs are omitted in 'This green.', verbs are substituted in 'I put them to the doctor's.', and verbs are used redundantly in 'He'll might get in jail.'

At the transformational level, having optionally chosen a transformation, the children sometimes do not observe all of the rules that are attendant on the use of this transformation. As an example, at the beginning of the age range when the verb phrase is being expanded from just the main verb to <u>be</u> + <u>present</u> <u>participle</u> + <u>verb</u> and the contraction transformation is applied, as it always is except in emphatic sentences, the children sometimes omit the contracted form and produce 'I going.' Later on, when the verb phrase is further expanded to <u>have</u> + <u>perfect</u> <u>participle</u> + <u>be</u> + <u>present</u> <u>participle</u> + <u>verb</u> and the contraction transformation is applied, the same result occurs and 'I been thinking about that.' is produced.

The most commonly used general transformation is conjunction, in which one sentence is added to another sentence. Sometimes the children use the conjunction transformation without using the rule that verb tense in the second sentence must agree with verb tense in the first sentence. An example of this is 'They mixed colors and pour buckets.'

At the morphology level of the grammar, again, omissions of rules, redundancies in rules, and substitutions take place. These occur with verbs, nouns, adverbs, pronouns, and possessive forms. As examples: noun endings are omitted in 'She has lots of necklace.' and 'He's next to a few stone.' Noun stems are substituted in 'Those are wolfs.' and 'We have childs in this school.', and noun endings are used redundantly in 'There's furnitures.' and 'Where are the peoples?'

Some quantitative statements can be made about the use of these restricted forms.

In general, the use of alternate rules gradually declines from the beginning of the age range to its end. There are significant differences found when one compares the percentages of children using these forms at the beginning of the age range and at its end. However, this decline is not asymptotic in nature, but, rather, fluctuating throughout the age range. Indeed, the specific structures that are formulated by alternate rules and the particular alternate rules used by a sizable number of the children at any given age period change as an increasingly mature population is observed.

When we divide the age range into 4-month periods and average the age periods in which there is peak usage of restricted forms, that is, all types of omissions, redundancies, and substitutions and all types of nonobservations of rules needed to produce simple and general transformations, and we then average the peak percentages of children using these forms, we see the following:

At the phrase-structure level the type of restricted form that peaks earliest and highest is omission. Then comes substitution, and finally redundancy. At the transformation level we see that nonuse of rules with simple transformations peaks earliest and highest. The peak for nonuse of rules with general transformations occurs later, when, in fact, more of the children are using some general transformations. At the morphology level, again, omissions peak earliest, then substitutions, and later, redundancies. At this level redundancies peak highest.

These trends can be seen in Figs. XVIII-1, XVIII-2, and XVIII-3. The real percentages of children using these forms at the beginning of the age range (from 2 years, 10 months to 3 years, 2 months) and at its end (from 6 years, 10 months to 7 years, 1 month) are also given.

I believe that there are qualitative, as well as quantitative, statements that can be made. The children's usage of grammar, and the word usage is stressed, did become increasingly complex over the age range observed. This complexity is not related simply to increasing sentence length or proportion of usage of what has been termed compound and complex sentences or, in the terms of this model of a grammar, general transformations. These changes seem to be extensions of behavior without additional rules in the grammar. Conjoining two sentences needs the same application of rules as conjoining three or four. To delete and substitute as in relative clauses needs the same application of rules whether we do it once in ten minutes or three times. This increasing complexity also seems to be very dependent on the child's ability to proceed from the application of the most general rule to the application of increasingly differentiating rules to produce a particular syntactic structure. In accordance with the model used, all instances of the use of a restricted form represent the use of an elementary rule, without or with some of the additional steps required. That is, rules for addition, deletion, permutation, and substitution were applied but without observation of ordering in some instances and in other instances without using the combination of these
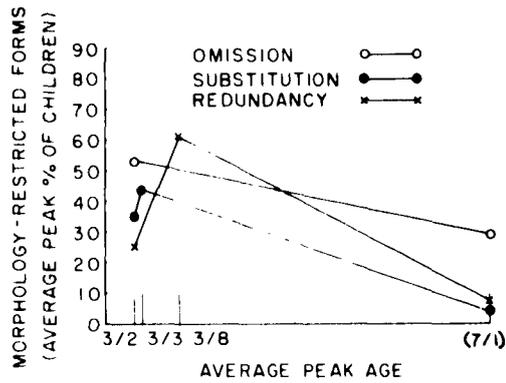
Fig. XVIII-1. Ages for peak usage of restricted forms at the phrase-structure level.
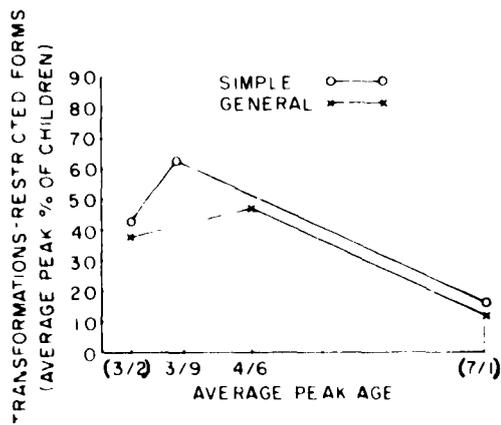
Fig. XVIII-2. Ages for peak usage of restricted forms at the transformation level.
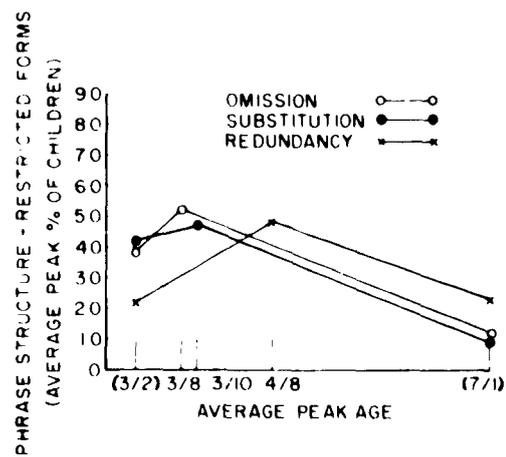
Fig. XVIII-3. Ages for peak usage of restricted forms at the morphology level.

elementaries required to produce the completed form of these structures.

Varying degrees of generalization take place in this process from greatest gener-
alization to increasing differentiation to complete differentiation and, I believe, new
organization. This process seems to reflect the hypothesized model, that is, the child
determining by a series of successive approximations the rules used to generate a sen-
tence. For example, omission would represent the application of the most general rule,
then substitution, and redundancy before complete differentiation. One aspect of this
new organization is differentiation between nonterminal and terminal rules, for example,
to formulate that the use of a modal, such as can, may or will, + verb is a terminal rule
and that two modals are mutually exclusive in the same context. We may say, 'He might
get into jail.' or 'He will get into jail.' but not 'He'll might get in jail.' Other examples
are to formulate that the question transformation is terminal and cannot be substituted
to produce new sentences (we say, 'What is that?' but not 'I know what is that.'), and
that certain endings of verbs and nouns are also terminal. We say pushed and people,
but not pushted and peoples.

Perhaps we can say that younger children's usage of grammar is simpler than older
children's or adults' usage of grammar because proportionally more of the younger
children use an incomplete set of rules to produce some syntactic structures and because
increasing levels of differentiation in the use of rules are found, going toward complete
differentiation, as older children are observed. The word usage is again stressed
because, although in some instances a child applies the elementary rule or an incom-
plete set in the formulation of a syntactic structure, in other instances he applies the
complete set of ordered rules to the formulation of this same structure. He seems to
display competence although this is not always realized in performance.

We now come to the question of what these data mean in terms of an hypothesis of
motivated imitation in language acquisition. The limitations of the nervous system for
memorizing all instances of sentences heard and storing them for later use seem to obvi-
ously negate the hypothesis of language acquisition as primarily an imitative function.
However, there are some theorists who divide language learning into a twofold process:
Early learning is the establishment of stimulus response associations with responses
learned by imitation, and later learning (the first meaningful utterances) is accomplished
by cognitive processes. From this point of view one would predict that the sounds and
words produced first would represent complete mastery of articulation, but this is not
the case. Children produce sentences long before articulation has been mastered. Finally,
if language acquisition is an imitative function, then children should be producing,
first, sentences with omissions because of the limitations of memory, and then complete
sentences. One might assume that the other types of restricted forms produced by chil-
dren are a result of imitation of peers. In that case one would expect a very random
production of these restricted forms. The results of this study seem to indicate that the

process is neither random nor one of remembering to put in more of the missing parts of sentences as the child matures. Rather, the restricted forms produced beyond omission, such as substitution and redundancy, reflect the child's improved ability to generate particular structures from increasingly more differentiated sets of rules as he matures.

Paula Menyuk

References

1. B. F. Skinner, Verbal Behavior (Appleton-Century-Crofts, Inc., New York, 1957).

2. J. Dollard and N. E. Miller, Personality and Psychotherapy (McGraw-Hill Book Company, Inc., New York, 1950).

3. O. C. Irwin, Infant speech, J. Speech and Hearing Disord. 16, 159-162 (1951).

4. D. McCarthy, Language Development of the Preschool Child (University of Minnesota Press, Minneapolis, 1930).

5. M. Halle and K. Stevens, Speech recognition: A model and a program for research, Trans. IRE, Vol. IT-8, pp. 155-159, February 1962.

6. G. H. Matthews, Analysis by Synthesis of Natural Languages, Internal Memorandum, Mechanical Translation Group, Research Laboratory of Electronics, M. I. T., 1961.

7. G. A. Miller and N. Chomsky, Finitary models of language users, Handbook of Mathematical Psychology, Vol. 2, edited by R. R. Bush, E. H. Galanter, and R. D. Luce (John Wiley and Sons, Inc., New York, in press).

8. N. Chomsky, Syntactic Structures (Mouton and Company, 's-Gravenhage, 1957).

9. P. Menyuk, Syntactic rules used by children from preschool through first grade, J. Child Development (in press).

XIX. ADVANCED COMPUTATION RESEARCH[*]

Prof. H. M. Teager   U. Shimony   R. L. Ward
E. L. Ivie   D. U. Wilde

A. NETWORK ANALYSIS BY DIGITAL COMPUTER

The amount of work required to analyze an electrical network grows rapidly as a function of the number of network nodes, N. The digital computer is a tool that can reduce this growth rate if the analysis is done in a proper manner. A study of analysis procedures that are applicable to digital-computer automation can be divided into two parts: topological and matrix. Such a study shows that the computation time required by topological methods grows as N!, while that of matrix methods can be made to grow more slowly. The use of matrix methods requires a set of independent Kirchhoff voltage or current equations. Further evaluation shows that node-to-datum voltage equations require a minimal amount of equation setup time. These equations are linear in the node-to-datum voltage variables but have polynomial coefficients of the form:

$$Y_{ij} = C_{ij}D^{+1} + G_{ij}D^{+0} + K_{ij}D^{-1},$$

where the C's, G's, and K's are either numerical or symbolic literals, and the D's are either time-domain operators or frequency functions. Because of this generality of coefficient and because most practical electrical networks have few interconnected node pairs, Cramer's Rule was chosen as the procedure to solve these equations. Cramer's Rule requires the expansion of two determinants.

The work for expanding a determinant by the straightforward application of the classical Laplace method grows factorially. The topology of the determinant expansion reveals a treelike structure, referred to as the determinant tree.

$$\Delta = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$$

$$= a(ei-fh) - b(di-fg) + c(dh-eg)$$

The Determinant
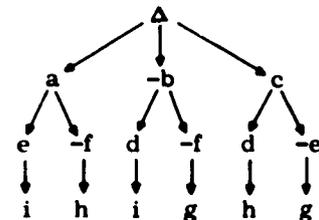
The Determinant Tree

If the tree is pruned from bottom to top, the order of the simplification is the same as that of evaluating the determinantal equation, starting on the inside of the parentheses and working outward. Thus, the determinant tree indicates the minimum expansion

---

work procedure. The structure of the tree can be pruned as the expansion progresses, and the memory space required to store the tree grows only as N^3. The work for expanding the determinant of a general network by the tree algorithm grows as $(N-1)!$. However, the construction rules of the determinant tree allow the computer to greatly reduce the expansion work for networks having few interconnected node pairs. For example, experimental results show that the work required for analyzing a ladder network grows geometrically at approximately 1.5 per node instead of factorially.

A program has been written in the MAD language and tested on sample networks. The input is a list of network elements and their terminal nodes, the desired unknown voltage, and the specified datum node. The present state of the loading routine allows for linear, passive nonmutual elements and for independent, current-controlled, and voltage-controlled current sources. Also, one symbolic circuit element may be included to allow the study of the circuit behavior as a function of this element. The output of the program is the desired unknown voltage as a ratio of polynomials. By using a unity current input, the input admittance or any transfer admittance may be found. The loading routine can easily be modified to handle any other linear circuit elements and more symbolic elements. Also, the expansion subroutine can be generalized for use with determinants with general polynomial entries.

<div align="right">D. U. Wilde</div>

XX. COMMUNICATIONS BIOPHYSICS[*]

Prof. W. A. Rosenblith[†]
Prof. M. A. B. Brazier[‡]
Prof. M. Eden
Prof. M. H. Goldstein, Jr.
Prof. W. T. Peake
Prof. W. M. Siebert
Dr. J. S. Barlow[**]
Dr. A. Cavaggioni[††]
W. A. Clark[†]
Dr. B. G. Farley[†]
Dr. G. L. Gerstein[†]
Dr. E. Giberman[‡‡]
Dr. R. D. Hall
Dr. N. Y-S. Kiang[***]

Dr. T. T. Sandel[†]
Dr. Eda Berger Vidale
J. A. Aldrich
R. M. Brown
S. K. Burns
R. R. Capranica
Eleanor K. Chance
R. J. Clayton
A. H. Crist
P. R. Gray
J. L. Hall II
F. T. Hambrecht
J. G. Krishnayya
R. G. Mark

P. Mermelstein
C. E. Molnar[†††]
Ann O'Rourke
R. R. Pfeiffer[***]
Cynthia Pyle
M. Sachs
D. M. Snodderly, Jr.
G. F. Svihula
I. Thomae
T. F. Weiss
J. R. Welch
M. L. Wiederhold
G. R. Wilde
S. Zilles

R. C. Baecker
B. O. Butterfield
T. Cohn
C. W. Einolf
A. Fevrier

A. Gabrielian
E. J. Hoffman
D. K. Joseph
D. Langbein

W. C. Marmon
R. M. Mason
J. Meyer
J. A. Parnell
A. P. Tripp

A. MINIMUM DETECTABLE INTERAURAL INTENSITY DIFFERENCE

We have measured the precision of listeners in judging the relative position of a lateralized binaural stimulus with respect to a centered binaural stimulus. The stimuli were clicks produced by applying 100-μsec rectangular voltage pulses to PDR-10 earphones. The centered stimulus was a train of 15 clicks at a rate of 10 per second with zero interaural intensity difference. The lateralized stimulus was a similar train varied in interaural intensity difference. The centered and lateralized stimuli were presented at the same average intensity, and this average intensity was a parameter of the experiment.

Figure XX-1 shows the performance of listeners in this task as per cent correct lateralizations versus interaural intensity difference for the lateralized stimulus, with

Fig. XX-1. Per cent correct lateralizations vs interaural
intensity difference. (Average of 3 subjects.)

average intensity as a parameter. These results are inconclusive because they repre-
sent the average over three subjects, and there was considerable intersubject varia-
bility. The 75 per cent correct level lies between 0.4 db and 0.6 db, and there is some
indication that the 75 per cent correct level increases as average intensity increases.

D. K. Joseph, J. L. Hall II

B. LATERALIZATION PERFORMANCE FOR SINGLE CLICKS BY
THE METHOD OF PAIRED COMPARISONS

We have measured the precision of listeners in judging the relative position of a
lateralized binaural stimulus with respect to a centered binaural stimulus. The stimuli
were clicks produced by applying 100-μsec rectangular voltage pulses to PDR-10 ear-
phones. The centered stimulus was presented with zero interaural time disparity; the
lateralized pair varied in interaural time disparity. This stimulus configuration affords
a minimum of information and/or redundancy to the listener and can be considered a
limiting case for comparable configurations of stimuli in similar experiments, given
a paired comparison task. A more general treatment would include differences in num-
ber of clicks presented at various rates for both the standard and test sequences.

Figure XX-2 shows the performance of listeners in this task as per cent correct
lateralization versus interaural delay for the lateralized pair. For comparison, the
performance of two listeners on a highly redundant stimulus sequence is also shown.
The redundant sequence consisted of a 2-sec click train at 25 clicks per second for the
centered stimulus, and a 2-sec click train at 5 clicks per second for the lateralized
stimulus. The data for redundant stimuli are from experiments that are being performed
at present and will not be discussed further in this report, except to note that the per-
formance is much better with such signals than it is with single click pairs.

These results lead us to conclude that this approach is potentially fruitful for under-
standing the nature of what might be considered optimal redundancy in binaural auditory
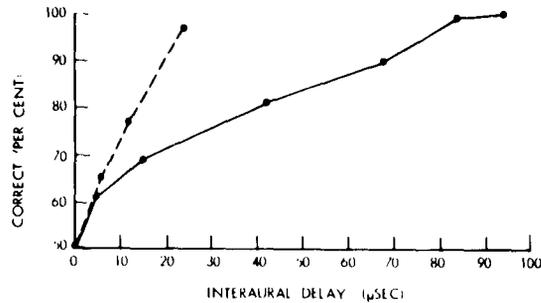
Fig. XX-2.

Per cent correct vs interaural delay. (Separation time, 0.5 sec.) Solid curve: Binaural clicks for 6 subjects. Dashed curve: Binaural click trains (50 in 2 seconds; then 10 in 2 seconds) for 2 subjects (Sandel and Hall).

lateralization performance, and the experiments in this direction will continue.

T. E. Cohn, J. L. Hall II, T. T. Sandel

C. EFFECTS OF OLIVOCOCHLEAR BUNDLE STIMULATION ON ACOUSTICALLY EVOKED POTENTIALS

Galambos[1] first showed that electrical excitation of the crossed olivocochlear bundle could reduce the amplitude of auditory nerve responses to acoustic stimuli. Recently, Desmedt,[2] recording with gross electrodes, and Fex,[3] recording with microelectrodes, have added quantitative information on the activity of this efferent auditory system. While the work reported here was in progress Desmedt[2] published results showing that olivocochlear-bundle (OCB) activation could be thought of as producing an equivalent decibel reduction of the acoustic stimulus, at least over an intensity range of 40 db from physiological threshold for gross-electrode evoked responses.

In our experiments on cats, the floor of the fourth ventricle was exposed and stimulating electrodes were placed on the site of the decussation of the crossed OCB. Responses evoked by clicks and short noise bursts were recorded at the round window and from the skull over the auditory area of the cerebral cortex. We desired to record from "both ends" of the auditory system in order to determine whether the effects of OCB stimulation could be localized to the periphery.

The cochlear microphonic (CM) and the amplitude of the neural components N_1 and N_2 of round window recordings were measured, as well as the latency of N_1 and the response amplitude at the auditory cortex (C). Acoustic stimuli were presented at a rate of one per second and could be preceded by a tetanic train of shocks to the OCB.

In most respects the results obtained confirmed those that have been reported by Galambos[1] and Desmedt.[2] The optimum shock repetition rate for reduction of neural responses was found to vary from 330 per second to 400 per second. An optimum duration of the tetanus was found for a given repetition rate and a given delay from the end

of the tetanus to the acoustic stimulus. Too few shocks have no effect, and as the duration of the tetanus is made too long the reduction of response amplitude decreases. The optimum shock configuration was usually found to consist of approximately 60 shocks.

For stimuli within 25 db of VDL threshold the response amplitude could be reduced to the point where no response was seen in the average of 32 responses. When responses were reduced by increasing the strength of the shocks to the OCB the response component amplitudes were found to maintain very nearly the same relationship to one another as in the no-shock condition.

As the neural responses are reduced by OCB stimulation the amplitude of the cochlear microphonic is increased. In some cases, for relatively strong shocks, CM was doubled. As N_1 is decreased at a given click intensity, its peak latency does not maintain the same relationship to its amplitude as is found in a standard intensity series. Without OCB shocks the latency of N_1 decreases almost linearly with increase of N_1. At intensities within 20 db of VDL threshold the latency of N_1 increases as its amplitude is decreased by OCB stimulation, but at intensities ~40 db above threshold the latency remains constant as N_1 is decreased, while at higher intensities the latency may actually decrease. As the delay of presentation of the click from the end of the shock tetanus is increased, the N_1 response amplitude recovers to its no-shock value within 75-100 msec.

Desmedt suggests that a given shock configuration produces an effective reduction of acoustic stimuli that is "largely independent of the intensity which happens to be chosen for the testing click, provided the latter is kept within about 40 db of threshold."[4] If the effectiveness of OCB stimulation is measured as a percentage reduction of N_1, the reduction for a given shock configuration is seen to be much greater at low intensities than high (Fig. XX-3a). If the reduced response amplitude is matched to the response of a lower intensity click without OCB shocks, an equivalent decibel reduction in acoustic stimulus can be computed. Even with this measure, OCB shocks are seen to be somewhat more effective in reducing responses to low intensity stimuli (Fig. XX-3b).

To investigate the effects of OCB stimulation on high intensity stimuli, 0.1-msec noise bursts were used. With these stimuli the CM is a random function so that its amplitude in the average of 64 responses is small compared with N_1 and N_2. Without averaging, at intensities higher than 60 db above threshold, CM is so large that it often obscures N_1 and N_2 recorded at the round window. In Fig. XX-4 response amplitudes are plotted against intensity for noise-burst stimulation. It will be seen that there is a dip in the N_1 and N_2 intensity series at -30 db. When shocks are administered preceding a noise burst at an intensity
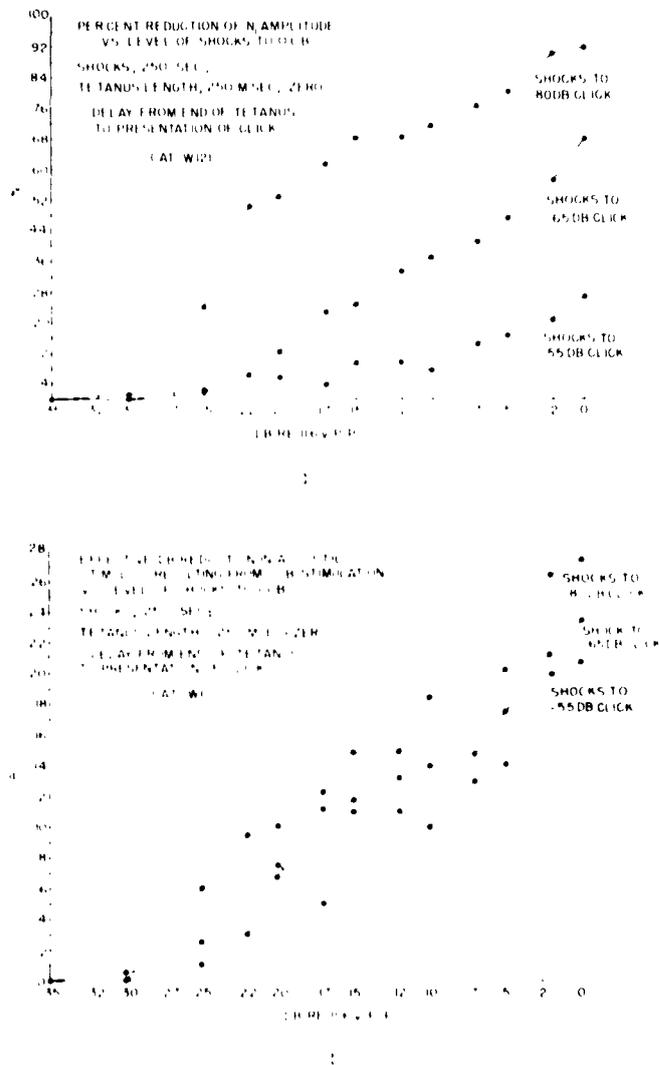
Fig. XX-3. (a) and (b) contain data from the same experiment. In both plots the abscissa gives peak-to-peak amplitude of shocks in db re 11.6 volts. Click intensities are in db re 5-volt 0.1-msec square pulse to a PDR-10 earphone. In (b) effective reduction is computed by matching reduced peak-to-peak N_1 amplitude with OCB shocks to no-shock conditions at a lower intensity. All points are from the averaged response to 32 identical stimuli.
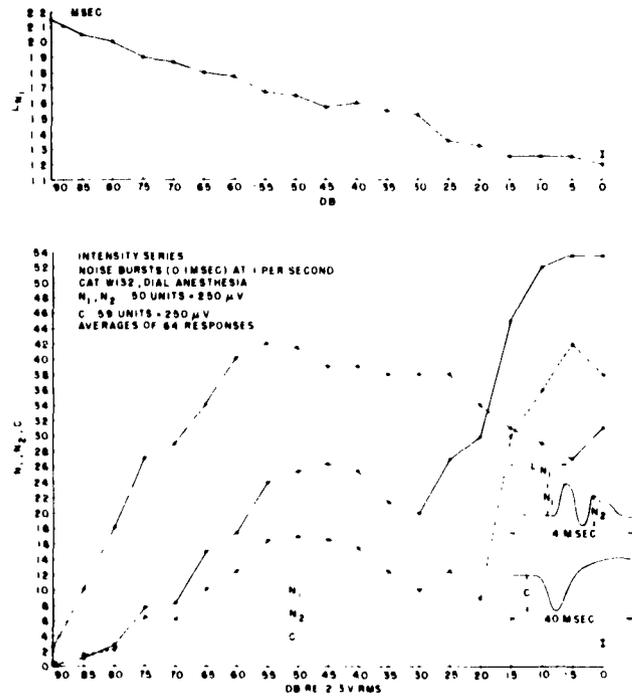
Fig. XX-4. Noise-burst intensity series. Noise bursts are produced by presenting noise of 30 cps-20 kc bandwidth to an electronic switch set to give 0.1-msec bursts. Rise time of switch, 25 μsec; 0 db noise is 2.3 volts rms.
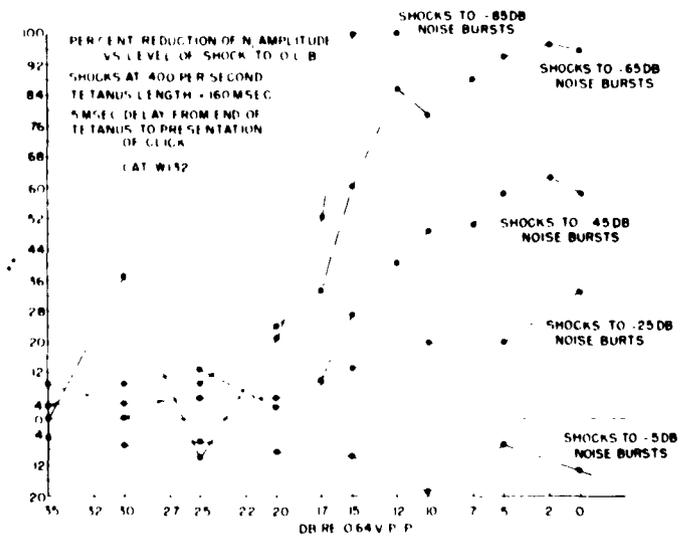
Fig. XX-5. Percentage reduction of N_1 base line-to-peak amplitude brought about by OCB shocks plotted as a function of the peak-to-peak voltage of shocks in db re 0.64 volt.

corresponding to the bottom of such a dip the response amplitude decreases, so that here the concept of an effective reduction of acoustic stimulus does not apply.

Figure XX-5 shows the percentage reduction (brought about by OCB stimulation) of the amplitude of N_1 response to 5 different intensities of noise bursts. The points for shocks below -20 db re 0.64 volt peak-to-peak give an indication of the fluctuations found in the averaged responses. Note that there is no evidence for reduction of the N_1 response to -5 db noise bursts, even with shocks 15 db more intense than those capable of completely abolishing the averaged response to -85 db noise burst.

Further details will be found in a Master's thesis of one of the authors.[5]

<div align="right">M. L. Wiederhold, Eleanor K. Chance</div>

References

1. R. Galambos, Suppression of auditory nerve activity by stimulation of efferent fibers to the cochlea, J. Neurophysiol. 19, 424 (1956).

2. J. E. Desmedt, Auditory-evoked potentials from cochlea to cortex as influenced by activation of the efferent olivo-cochlea bundle, J. Acoust. Soc. Am. 34, 1478 (1962).

3. J. Fex, Auditory activity in centrifugal and centripetal cochlear fibres in cat, Acta. Physiol. Scand. 55, Supp. 189 (1962).

4. J. E. Desmedt, op. cit., p. 1483.

5. M. L. Wiederhold, Effects of Efferent Pathways on Acoustic Evoked Responses in the Auditory Nervous System, S. M. Thesis, Department of Electrical Engineering, M. I. T., June 1963.

D. NOTE ON THE MACH ILLUSION

Under appropriate viewing conditions, the illusion discussed in this report appears to "invert." This inversion persists if object and/or observer move. We note that a possible explanation of this behavior is the utilization of "defocus" as a monocular depth cue.

The illusion[1] can be constructed by folding a 3 X 5 index card as in Fig. XX-6. An observer looking into the pleats monocularly can easily convince himself that near edges are far, and vice versa. As shown in Fig. XX-7, the projection that the observer
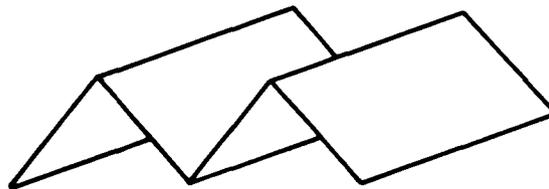
Fig. XX-6. The illusion.
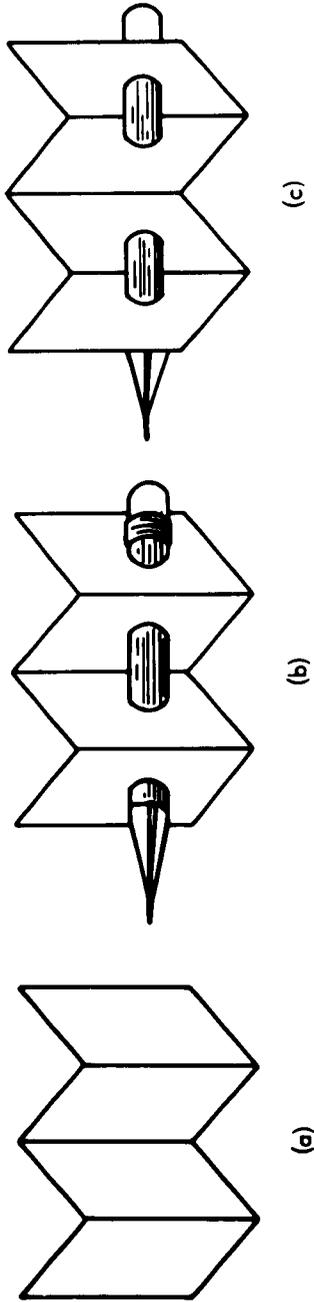
Fig. XX-7. (a) Front view of the illusion. (b) and (c) Two possible ways of interpreting the view shown in (a).
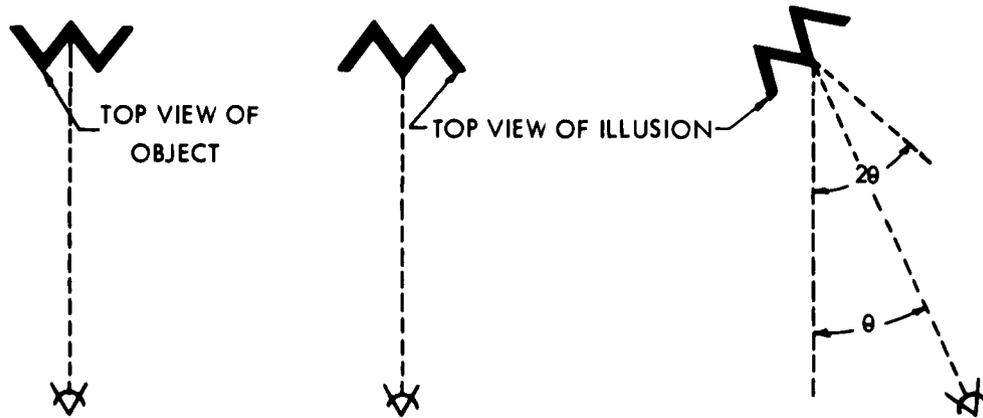
Fig. XX-8. As the observer moves, the illusion appears to move.

NORMAL TO PLANE ①, ②

(a)   (b)
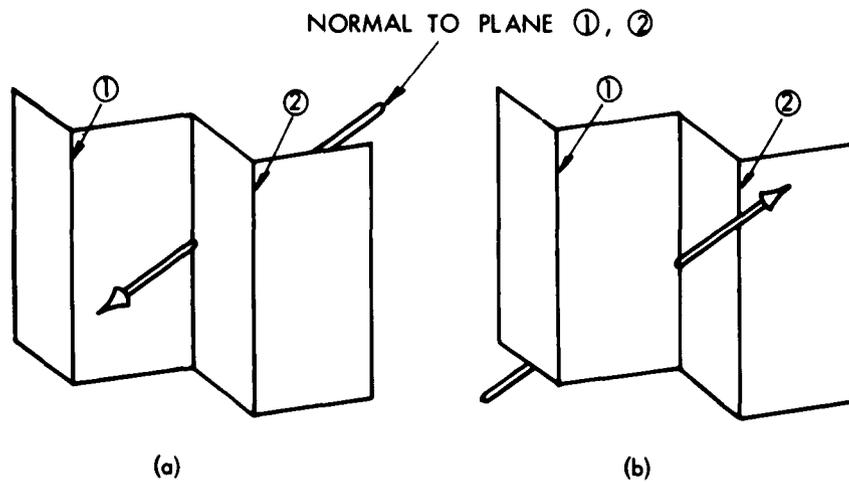
Fig. XX-9. (a) The object as viewed by an eye to the right of the normal to the plane
determined by lines ① and ② . The apparent position (b) of this normal
in the illusion is to the observer's right, thereby indicating that the illu-
sion has twisted more than the eye has moved.
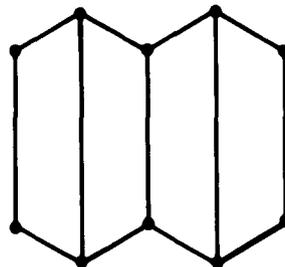
Fig. XX-10. Front view of card. The faces of the inversion are trapezoidal
with the small edges near the observer.

sees really is ambiguous, which accounts for the ease of inversion.

Unlike many similar illusions, the object is solid. Hence observer motion trans-mits new information to the eye. Even so, the illusion does not vanish — it appears to gyrate as the observer does. As Professor Eden[1] points out, the apparent motion is predictable by the postulate that the observer sees the reflection of the object in a mir-ror passing through the object normal to the line of sight. (See Figs. XX-8 and XX-9.) This approximation is invalid when the observer is close to the object.

When the eye is close to the object, distortions occur and the inversion is no longer the mirror image of the original (see Fig. XX-10).

As is apparent from the figures, many objects can produce the same projection on the observer's picture plane. Thus it is not remarkable that an illusion is possible. It is the fact that only one illusion configuration occurs that is remarkable.[2] See Fig. XX-11. The unique "choice" of the inversion from the set of all configurations
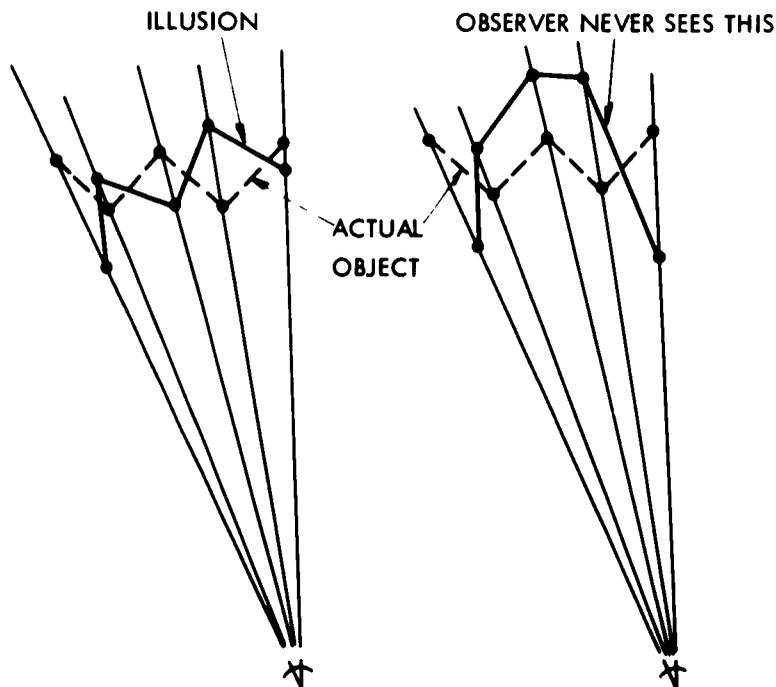
Fig. XX-11. The observer sees either the object or its inversion. He does not see any of the other configurations that make the same projection on the retina.

with identical retinal projections is not the result of:

A preference for rectilinearity. Curved surfaces are also uniquely invertible.

A preference for apparent stationarity. The illusion moves as the observer does.[3]

A preference for rigid body motion. The illusion distorts as it moves.

Overlap cues.

Shading. Shadows are often nonphysical in that no light source could generate them.

Texture gradients. All of these gradients are backwards in the illusion.

If we postulate that the eye is sensitive to the blur resulting from slight amounts of defocus, this singular inversion can be readily explained. The circle of confusion created by a point 31 cm from an eye focussed at 30 cm is approximately 2 microns across, roughly the distance between cones in the fovea, or the diameter of the diffraction pattern of the pupil. Acuity tests indicate that in some situations effects as small as this are readily perceived.

As shown in Fig. XX-12, when the eye is focussed at one range, points nearer to the eye can produce the same blur as points farther away. With the thin-lens approximation, $1/x + 1/z = 2/y$.
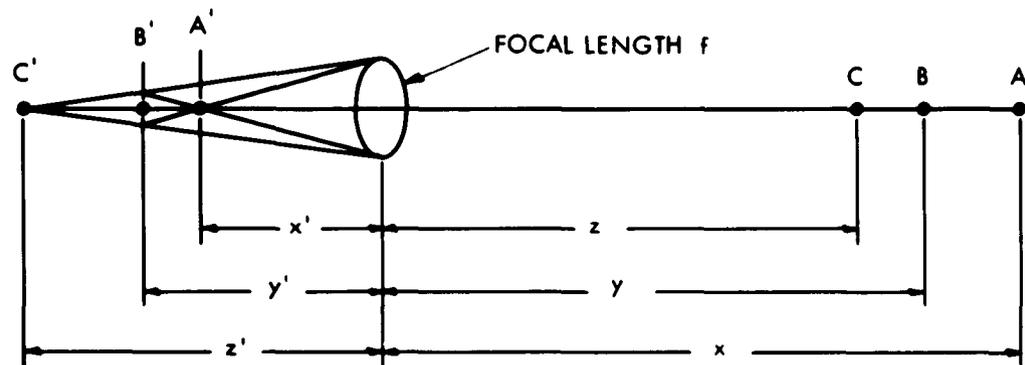
Fig. XX-12. The screen images point B. Points A and C produce the same circle of confusion on the retina. From the lens formula,

$$\frac{1}{x'} + \frac{1}{x} = \frac{1}{f}$$

$$\frac{1}{z'} + \frac{1}{z} = \frac{1}{f} \qquad \text{which imply} \quad \left(\frac{1}{x'} + \frac{1}{z'} - \frac{2}{y'}\right) + \left(\frac{1}{x} + \frac{1}{z} - \frac{2}{y}\right) = 0.$$

$$\frac{1}{y'} + \frac{1}{y} = \frac{1}{f}$$

By similar triangles,

$$\frac{y' - x'}{x'} = \frac{z' - y'}{z'} \qquad \text{which implies} \quad \frac{1}{x'} + \frac{1}{z'} - \frac{2}{y'} = 0.$$

Thus $\dfrac{1}{x} + \dfrac{1}{z} = \dfrac{2}{y}$.

Thus two points can be mistaken for one another if they lie on the same radial to the eye and the harmonic mean of their distances from the eye is the fixation distance.
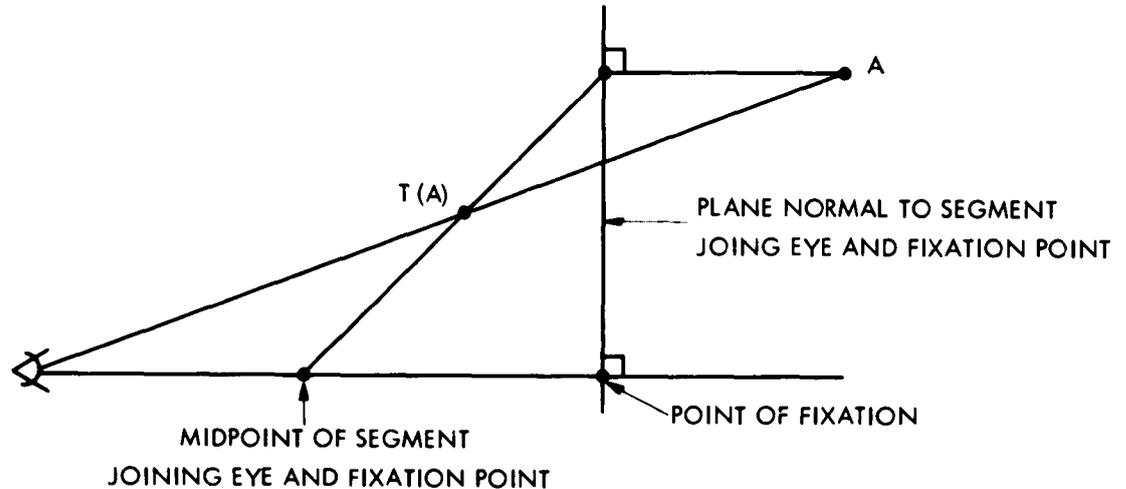
Fig. XX-13. A construction for T.

The transformation, T, mapping each point into the unique other point that produces the same defocus blur is easily visualized with the construction of Fig. XX-13. Some properties of T are:

T keeps straight lines straight.

$T^2 = I$.

At large fixation distances, T approaches a reflection.

Transformation T can be applied to objects and generally the resultant point set is in subjective agreement with the illusion.

Unfortunately, for objects subtending large angles at the eye, straight lines appear to be curved slightly. At best we could expect the thin lens formulas to be valid near the optical axes. Moreover, T shows a sensitivity to fixation distance which is difficult to observe in the illusion. Further tests are being made to estimate the magnitude of these deviations.

W. L. Black

References

1. See M. Eden, A three-dimensional optical illusion, Quarterly Progress Report No. 64, Research Laboratory of Electronics, M. I. T., January 15, 1962, p. 267 for a more complete description.

2. One can invert only part of an object, but the inverted portion is congruent to the corresponding portion when the whole object is inverted.

3. In an inverting telescope the inversion seems to be stationary and the object seems to move as the observer does. Such a telescope does not invert depth; it rotates the scene through π radians about the axis of the telescope.

E. WORK COMPLETED

These brief reports are abstracts of theses and a summary of a project completed during the past academic year, and not reported on previously.

Instrumentation

1. AN ON-LINE DIGITAL ELECTRONIC CORRELATOR

One of the many applications of correlation techniques has been their use in the analysis of waveforms recorded during electrophysiological experiments.

A suggestion of C. E. Molnar has led to the design of a digital correlator that can process signals with the frequency range of EEG data on-line. The correlator quantizes the input signal to 16 levels (4 bits). Samples of the input signal are taken periodically, the shortest sampling period being 5 msec. For this sampling rate the input signal must be filtered so that its spectrum is practically zero at higher frequencies than 100 cps. Autocorrelograms or crosscorrelograms are computed by a novel technique of precessing previously stored samples on a delay line relative to stationary accumulating sums. Each of these accumulating sums corresponds to the value of the correlogram at a particular value of delay. Samples emerging from the delay line are always displaced in time from the present input sample by an amount corresponding to the value of the associated accumulating sums.

The correlator has two modes of operation, one mode providing 200 correlogram points with an observation interval of several minutes, and the other providing 100 correlogram points with an observation interval up to several weeks. The sampling interval may be changed by a switch, with values of 5, 10, 20, 40, and 80 msec provided. This is also the interval between computed points in the correlogram. The input signal must be filtered so that it contains practically no energy at a frequency that is higher than the reciprocal of twice the sampling interval. Autocorrelograms are computed for positive delays only, whereas crosscorrelograms are computed for both positive and negative delays.

The correlogram is available as an analog signal from an 8-bit digital-to-analog converter, and is displayed continuously on an oscilloscope during computation. A strip chart recorder output is also provided. Digital readout on punched paper tape is possible with minor modifications.

The correlator is entirely composed of commercially available digital modules.

This study was presented to the Department of Electrical Engineering, M. I. T., in

partial fulfillment of the requirements for the degree of Master of Science, May 15, 1963.

G. R. Wilde

2. A MULTICHANNEL ELECTROENCEPHALOGRAPHIC TELEMETERING SYSTEM

This report is an abstract of an S. M. Thesis submitted to the Department of Electrical Engineering, M. I. T., May 17, 1963.

A four-channel electroencephalographic telemetering system that is small enough to be carried by an animal that is of the size of a cat has been designed and built.

Descriptions of the differential amplifiers, pulse duration multiplexing circuitry, transmitter and receiver, and decoding circuitry are given. Details for reproduction of the system, and a simple module method for adding more channels are described in the author's thesis.

Each channel of the four-channel system has a frequency range of 0.7-2000 cps with a noise level referred to the input of 5 microvolts peak-to-peak at 10,000 ohms. The over-all operating range of the system is 200 feet.

A report based on this thesis will appear as Technical Report 413 of the Research Laboratory of Electronics.

F. T. Hambrecht

3. A SIGNAL GENERATOR FOR USE IN FREQUENCY-DISCRIMINATION EXPERIMENTS

A self-contained test generator for use in frequency-discrimination measurements has been developed. A random-word generator is used to control the sequence of presentations to the subject; this generator may be set to give 0, 0.25, 0.5, 0.75, and 1.0 probabilities of a high-then-low frequency presentation. The total number of tests and decisions is counted by electromagnetic counters. Provision is made for time-chopping of the output signal so that its effects on the discrimination ability of the subject may be evaluated.

This study was submitted as a Bachelor's thesis to the Department of Electrical Engineering, M. I. T., 1963.

A. P. Tripp

4. LOW-FREQUENCY RATE METER WITH FAST RESPONSE

This instrument is a self-contained, completely transistorized, portable frequency meter that is capable of measuring rates from 3 cycles per minute to 500 cps. With the use of 12 ranges, an accuracy of ±5 per cent can be achieved. The response is such that any change in frequency between two input pulses will be detected.

The interval between each consecutive input pulse is measured and a voltage

proportional to this interval is generated. This voltage remains constant during the following interval. The pulse frequency is the reciprocal of this voltage.

This study was submitted as a Bachelor's thesis to the Department of Electrical Engineering, M. I. T., 1963.

C. W. Einolf, Jr.

Psychophysics

1. INTERAURAL TIME-DIFFERENCE THRESHOLDS FOR BANDLIMITED NOISE

In a large class of auditory discriminations, significant information is carried in the changing structure of the signal in time. This research concerned one way of studying the sensitivity of the auditory system to such time structure — the ability of human observers to discriminate interaural time shifts in bandlimited noise signals — and measurements were made by psychophysical methods. The psychophysical data can be viewed as reflecting the limitations imposed on performance by nonadditive signal-dependent noise in the auditory system. It was found that performance improves strongly with increasing bandwidth and signal energy, over the range of the parameters employed in this investigation. There was considerable evidence of the effect of non-sensory factors on performance.

This study was submitted to the Department of Electrical Engineering, M. I. T., 1963, in partial fulfillment of the requirements for the degree of Electrical Engineer.

J. A. Aldrich

2. HUMAN DISCRIMINATION OF AUDITORY DURATION

In a Bachelor's thesis submitted to the Department of Physics, M. I. T., 1963, a psychophysical model combining statistical decision theory with knowledge of the anatomy and physiology of audition is discussed. In the framework of this theory an experiment on the judgment of auditory duration was planned and executed. The model's predictions for the experiment are presented. Three subjects were tested in a forced-choice experiment that was perfomed to judge which of two auditory signals was longer. Their discrimination between 1000-cps sinusoids separated by empty intervals and empty intervals separated by sinusoids was investigated at three signal durations, 300 msec, 100 msec, and 33 msec.

The model predicts reasonably well the shape of the function relating percentage correct to increment duration. There are large individual differences in comparative performance at the three base durations. Interpretations of these effects are proposed and discussed. There appears to be a significant sensory difference between the judgments of the filled and the empty intervals. Suggested modifications of

experimental procedure are made in the light of the data and the prior results of other investigators.

R. C. Baecker

3. A PROCEDURE FOR n-ALTERNATIVE FORCED-CHOICE PSYCHOPHYSICAL TESTING

In a Bachelor's thesis submitted to the Department of Physics, M. I. T., May 17, 1963, an n-alternative forced-choice testing scheme is described. First, a brief explanation of the purpose of the study and the approach that was used throughout the experimental work is given. Second, a description is given of the device that was used in these tests, the design and construction of which comprised the major part of this work. Finally, data that were obtained are used to point out one of the pitfalls in using this testing scheme.

J. S. Meyer

Physiology

1. ACTIVITY OF SINGLE NEURONS IN THE LATERAL GENICULATE OF THE RAT

This report is an abstract of a thesis submitted to the Department of Electrical Engineering, M. I. T., June 1963, in partial fulfillment of the requirements for the degree of Master of Science.

This research is based on a preliminary investigation of the problem of describing the receptive fields of single cells in the lateral geniculate of lightly anesthetized rats. Patterns of ongoing single-unit activity exhibiting the "grouped characteristics" reported by previous experimenters are demonstrated and discussed. The effects of anesthesia are considered and variations in responsiveness to photic stimulation are noted.

Two types of receptive fields of the lateral geniculate of the rat are described. One is an "ON" unit, the other an "OFF" unit. Both have a uniform response type throughout the field when a small spot of light is used as a stimulus, and are strongly responsive to diffuse light. These two fields are compared with similar receptive fields found in other animals.

Tungsten microelectrodes were used and a procedure for manufacture of the electrodes, including a mechanized technique for etching, is described.

D. M. Snodderly, Jr.

2. ANALYSIS OF A CAT'S "MEOW"

The recorded "meows" from four cats were studied in this project. The principal conclusion is that in the majority of instances, the "meow" of the cat is determined

primarily by the rate of vocal-fold vibration as opposed to the shape of the vocal and nasal tracts.

J. A. Rome

3. VOICE OF THE BULLFROG

The vocalizations of the bullfrog (Rana catesbeiana) were studied by spectrographic analysis of recordings from three different sources, including laboratory recordings of male and female frogs. Oscilloscope pictures of the sound-pressure waveform were also studied. The vocal source is a train of pulses which has a repetition rate of approximately 90 per second. The spectral distribution of the call is primarily in two bands: 90-600 cps, and 900-1600 cps. A model of the bullfrog vocal system is proposed. Vocalizations of eight other Rana species are discussed.

This study was presented to the Department of Physics, M. I. T., as a Bachelor's thesis, 1963.

A. Fevrier

4. INVESTIGATION OF FILTER SYSTEMS FOR THE EKG WAVEFORM

In a Bachelor's thesis submitted to the Department of Electrical Engineering, M. I. T., 1963, the EKG waveforms of anesthetized frogs were studied to determine the optimum single-band filtering system for the EKG. The R pulse of the QRS complex was chosen as the strongest indicator of the heart beat. It was found that the signal-to-noise ratio for this pulse is optimum in the 15-30 cps frequency band.

W. C. Marmon

XXI. NEUROPHYSIOLOGY[*]

W. S. McCulloch	R. C. Gesteland	Diane Major
M. A. Arbib	W. L. Kilmer	L. M. Mendell
F. S. Axelrod	K. Kornacker	W. H. Pitts
P. O. Bishop	W. J. Lennon	J. A. Rojas
M. Blum	J. Y. Lettvin	A. Taub
J. E. Brown		P. D. Wall

A. COLOR VISION I

The range of visible light lies between 4000 Å and 7000 Å wavelength. Extremely powerful radiation above and below those limits can be seen, but these are the commonly accepted extrema. Within that band the spectral distribution of power in a ray of light describes some of its qualities. The color that is perceived seems to be related to that spectral distribution. Two of the operations that one can imagine to be made on the distribution are

1. Integration of power along the spectrum weighted by a function of position in the spectrum; and

2. A description of the distribution of power that is normalized with respect to total power, that is, the power in every region of the band divided by the power in the whole band.

The first operation takes (a) the "intensity" of light, and the second gives us the "quality." A suitable handling of the quality can be made to show (b) where power is greater along the spectrum, and (c) how much greater it is there than elsewhere. Our subjective impression for a uniform patch of light is that we see a color that has (a) brightness, (b) hue, and (c) saturation. These two triads of properties seem to be related.

Color perception can be studied in two aspects: the "aperture mode," and the "object mode." What I name as the color of a uniformly lit surface of a particular absorption spectrum depends greatly upon what other colors bound that surface, are in its neighborhood or have just preceded it in time on that surface. These effects are not negligible. Simultaneous and successive contrasts are involved in colored afterimages, colored shadows, color constancy, and all of those other visual phenomena studied in the mid-nineteenth century. But this topic of color in the real world — the "object mode" — is not discussed here; in this report I concern myself solely with what is called the "aperture mode," which is exemplified by this supposition: Suppose that my whole field of vision is a uniformly and neutrally colored matte surface in which there is a circular hole approximately 10° in visual angle. Let me partition that hole

into two semicircles and color the light coming through one uniformly. What manipulations with light are needed to change the other semicircle in such a way that I see a whole circle of uniform color? This method removes, as far as possible, the relation of color to form by rendering constant everything except color. Under such conditions, and for people with normal color vision looking directly at the hole, all colors can be matched by a combination of at least three colors; that is, three colors are sufficient, even though they may not be "real."

There are several principles that must be recalled before we continue.

1. Colors add to make a color.

2. A color cannot be dissected uniquely into component colors, and many different mixtures of different colors can give the same color.

3. Any color mixture, no matter how it is composed, must have the same appearance as the mixture of a certain saturated color (spectrum color or purple) with white.

4. Colors that look alike produce a mixture that looks like them.

5. When one of two kinds of light which are mixed together changes continuously, the appearance of the mixture changes continuously also.

(I recommend Chapter 20 in Helmholtz's Physiological Optics for further study.)

6. The laws of color mixture can be expressed as a barycentric operation on the spectral colors. If you think of the spectrum laid out along a wire, and have the wire bent round into an open plane figure, and if the power in a certain colored light everywhere along the spectrum is imitated by proportionate weights attached to the wire, then the center of gravity of the figure represents the resultant hue and saturation, and the weight of the whole figure represent the intensity of the color.

These are Newton's laws of color, and they have not been modified since his time. Grassman formalized these laws after Newton. Helmholtz gives a beautiful exposition of how they are used. Maxwell brought them to their modern form. The experimental problem is to erect the correct shape of the spectrum boundary of this figure, and Helmholtz shows how this can be done. The spectrum is not closed on itself as Newton had it, but is open just as a bent hairpin, and the colors, represented by mixtures of the extrema of the spectrum (the purples), lie along a straight line between the spectral extremes.

Once this barycentric method has been conceived, it is clear that, since any closed plane figure can be enclosed in a triangle, the same operation can be performed by using only three colors, one at each vertex of the triangle, with the power at any point in the spectrum considered a resultant of powers concentrated at those three points. Thomas Young announced this notion in its physiological form — that three types of retinal cell, each sensitive to a different color, were enough to describe a color space. This was later elaborated by Helmholtz.

But while Newton's laws ensure that the color plane is everywhere not concave, it

is everywhere convex over the locus of spectral colors, and straight only over the purples. Thus, of the colors needed to describe the color figure completely, two must lie outside the color figure, that is, they are imaginary. And, indeed, the reference colors in colorimetry are of this sort.

We are now able to continue our discussion of matching colors seen through the 10° aperture. Since I must use real lights and colors, I find that if there is a test color in one-half of the aperture and the three mixing lights are noncolinear (that is, one of them cannot be matched by a mixture of the other two), then I can either match the test color by an appropriate unique mixture of the other three or I can add one of the three to the test color and match a unique combination of these two by a unique combination of the other two. There is no set of real colors which I can choose for which, over all of the saturated colors to be matched, no color must be switched to the test side (that is, the intensity of that component is negative). Using four colors rather than three is no advantage, incidentally, for, since the locus of spectral colors is everywhere convex, at least two of the colors used for matching still will have to be imaginary, that is, they will have negative intensities for some of the saturated colors to be matched. The same is true for however many saturated colors that could be used to made a match; some of them will have to take on negative values in matching some other saturated colors. There is no advantage, then, in using more than three colors in colorimetry, and it is not as if there were a set of errors in the trichromatic scheme, which can be completely corrected by the use of one more color judiciously chosen.

"Color" is used here both as the subjective perception of light and as the physical properties of the light. Miss Ladd-Franklin was very severe with Helmholtz for doing this. I shall assume that the reader distinguishes "color" as a description of the intensity and quality of light from "color" as a description solely of the quality (when the intensity is being varied, as in color matching), from "color" as a subjective point in psychophysical space having intensity, hue and saturation, from "color" as the percept of hue and saturation independent of brightness. These distinctions are clear in context, I believe.

Having looked at color matching in the aperture mode and finding a trichromatic scheme sufficient, we now ought to consider the physiological question. Young, who knew about the discrete character of the retina — that it was a mosaic — (after all, Descartes based his theory of vision on the fact that the optic nerve fibers were separate elements), felt that it was unlikely that each fiber should have infinite modes of resonance, or the capacity to display simultaneously that which a spectrophotometer must do in time. Accordingly, he proposed, in fact, that the retina has three sorts of elements: one sensitive to red, another to blue, and a third to green; and that with these three types uniformly distributed color vision could be explained. This idea was taken up enthusiastically by Helmholtz, and the physiological consequence is the three-receptor hypothesis fitting with a trichromatic scheme of color matching. Another

physiologist, Hering, who was much taken by the notion of complementary colors, as implied by the barycentric system of Newton and measured by the nineteenth century physiologists, carefully pointed out that, since red was the complement of green and blue of yellow (as Democritus had remarked two millenia earlier), a three-duet system actually occurs physiologically. That is to say that we see color by virtue of taking an opposing relation between how much a "red" process is affected versus how much a "green" process is affected, so too "blue" vs "yellow" and "black" vs "white". The claim is advanced that this scheme accounts better for the facts of color blindness. A great deal of tedious nonsense cloaks this controversy. None of the evidence distinguishes the Hering from the Helmholtz notion, for, in fact, they are the same. To understand why, we must go back to the notion of colorimetry. The center of gravity of a triangle ABC, where the weights a, b, and c are concentrated at the apices, and the sum of the weights is constant, can be determined simply from the moments:

$$\frac{a}{a+b+c}\,AX, \qquad \frac{b}{a+b+c}\,BX, \qquad \frac{c}{a+b+c}\,CX,$$

where AX, BX, and CX are distances from each of the vertices to the center of gravity. If I use a fixed triangle, for which I know the distance of A to B, B to C, and C to A, then I also need to know the sum of all weights and two of the weights to determine where the center of gravity lies. Thus, if I know $\frac{a}{a+b+c}$ and $\frac{b}{a+b+c}$, it is sufficient for me to compute the center of gravity of a fixed triangle under the conditions given (that is, I need only to hang a triangle from two vertices to find the center of gravity). In fact, this is how modern colorimetry is carried out. We have all of the coupling here that we need. For example, if c is kept constant and a increases, then b decreases, and conversely, and all of this could be displayed entirely simply by one of the terms. The trichromatic scheme, when translated into a trichromatic physiological process, implies two things: First, only two of the three possible channels are required to provide all of the information about the color plane; and second, each of the colors denoted by each of the two channels will appear to be see-sawing with or opposing its complement. Hering's hypothesis, in the best light, modifies the Helmholtz hypothesis in this way: It is true that three colors are needed to erect the color space, but only two channels (that is, hue and saturation normalized against brightness) of output are required to carry information about the color plane. Total brightness, however, requires another channel.

When I speak about channels I imply that the processing of color data must occur very early in the visual system. There are several reasons for supposing this. First, there is an enormous dynamic range in our vision, and the reliable dynamic range of firing nerve fibers is not impressive. Second, any transformation of a continous

measure through a set of coupled nonlinear oscillators is bound to be noisy and it would seem a bit silly of nature to arrange for a noisy transmission only to recover the required signal by some sort of correlation process later; particularly, if the required signal can be achieved by direct methods earlier. Third, physiological studies on the optic nerve show that some such processing already occurs in the retina. Only the last reason is perfectly valid.

Although originally I planned to expound my particular theory of how color vision can be explained by form-function relations in the retina, I have now decided that this is premature. Those notions led me to the construction of a device to imitate color vision — but, since it could have been built purely by the laws of colorimetry, it proves nothing about my hypothesis.

Notice that the formula for determining the center of gravity is quite the same as

$$POS. 1 = -\frac{g_G}{g_G + g_R + g_B} V$$

$$POS. 2 = +\frac{g_R}{g_G + g_R + g_B} V$$
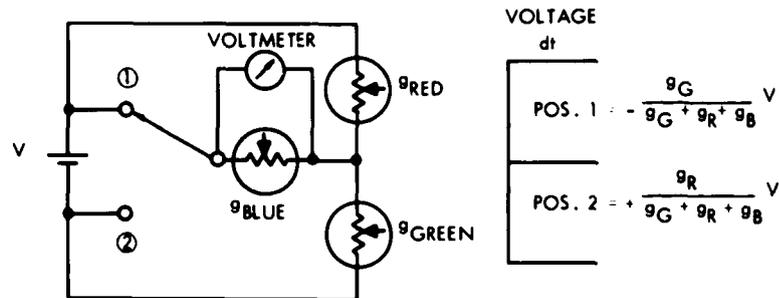
ALL PHOTORESISTORS SHOW LINEAR RELATION OF CONDUCTANCE TO THE POWER OF AN INCIDENT LIGHT OF FIXED SPECTRAL COMPOSITION

g_{RED} CONDUCTANCE OF PHOTORESISTOR OVERLAID WITH FILTER SO THAT CONDUCANCE AS A FUNCTION OF WAVELENGTH FOR EQUAL ENERGY ILLUMINATION IS CURVE 1

g_{GREEN} SAME AS CURVE 2

g_{BLUE} SAME AS CURVE 3
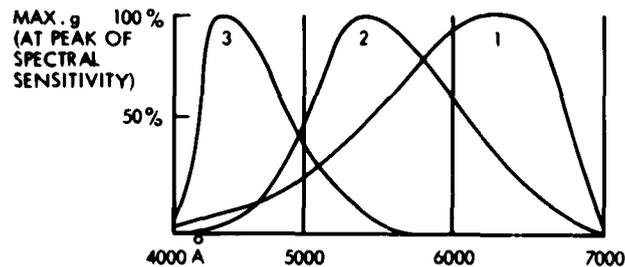
Fig. XXI-1. The Maggie.

Kirchhoff's laws for determining voltage at the conjunction of resistors. For example, I assume photoresistors (PR) whose conductance is a linear function of the intensity of light playing on them, and which have a flat response in the visible band. That is, I can write the conductance of one filter-PR combination as a function of its spectral sensitivity, for example, g_R. If the switches are in the up position as shown in Fig. XXI-1 and all PR's have identical characteristics, then, if we call the amount of light passed by the red filter R, by the green, G, and by the blue, B, then E becomes $E_R = \dfrac{g_R}{g_R + g_B + g_G} V$. If the switch is in the down position, $E = E_G = \dfrac{g_G}{g_R + g_B + g_G} V$. If E_R is taken to be the x-coordinate and E_G as the y, the locus of all "colors" that can be seen by this system lies within the triangle bounded by $(0, 0)$ $(0, V)$, and $(V, 0)$. Furthermore, a particular spectral composition of light playing simultaneously on all three elements uniquely determines a point within this triangle, independently of the intensity of the light. This is a very simple system and is useful for explaining the Newtonian laws of color to students. They see very quickly how the "opponents" theory of Hering and the trichromatic theory of Helmholtz are identical, and the switch between g_R and g_B is useful to illustrate this likeness.

But to make a practical commercial device from this is a bit more difficult. For one thing, while there are several linear photoresistive materials, some of them have nasty troughs in the visible spectrum (for example, in Clairex No. 2 material), and others that seem smooth on the linear plot of sensitivity vs spectrum but are somewhat insensitive in the visible band turn out to be very irregular in their spectral response (for example, in Clairex No. 4). This is why I dislike using linear rather than logarithmic plots of sensitivity vs spectrum. In general, the Clairex No. 5 material is quite good, and, despite its irregularity in the yellow, is essentially of a monotonic fall-off from its peak at 5600 Å.

If the sensitivity of the material at each point P in the spectrum is S_p, then the problem is to devise a filter with the property that if it attenuates the power at P by a factor, Q_p, the response of the material covered by the filter, $Q_p S_p$, will give the wanted sensitivity characteristic over all p's.

In order to build a usable color space, I do not believe that it is essential to work slavishly toward an exact imitation of the colorimetic X, Y, and Z. Preliminary study has convinced me that quite good color matches can be made by the crude use of Wratten filters with the Clairex No. 5 material. What is required of a combined filter and photoresistor is that it have a broadband, single-peaked sensitivity curve, falling off smoothly and monotonically away from the peak in both directions. It is important the the spectral colors be represented by three lines of numbers from the three filter-photoresistor combinations that vary smoothly and systematically from one end of the spectrum to the other. Thus the spectral order is preserved along the boundary of the color space, and

the space itself is linear and homogenous and arises barycentrically from the spectrum. Two facts are immediately apparent. First, whatever the other properties of this space are, it is a continuous map of human color space. There can be no discontinuities or crossovers. Thus if a color becomes redder to us, it will also become redder to the device. Second, it is directly distortable into many different projections of itself simply by adding true neutral density filters to the color filters.
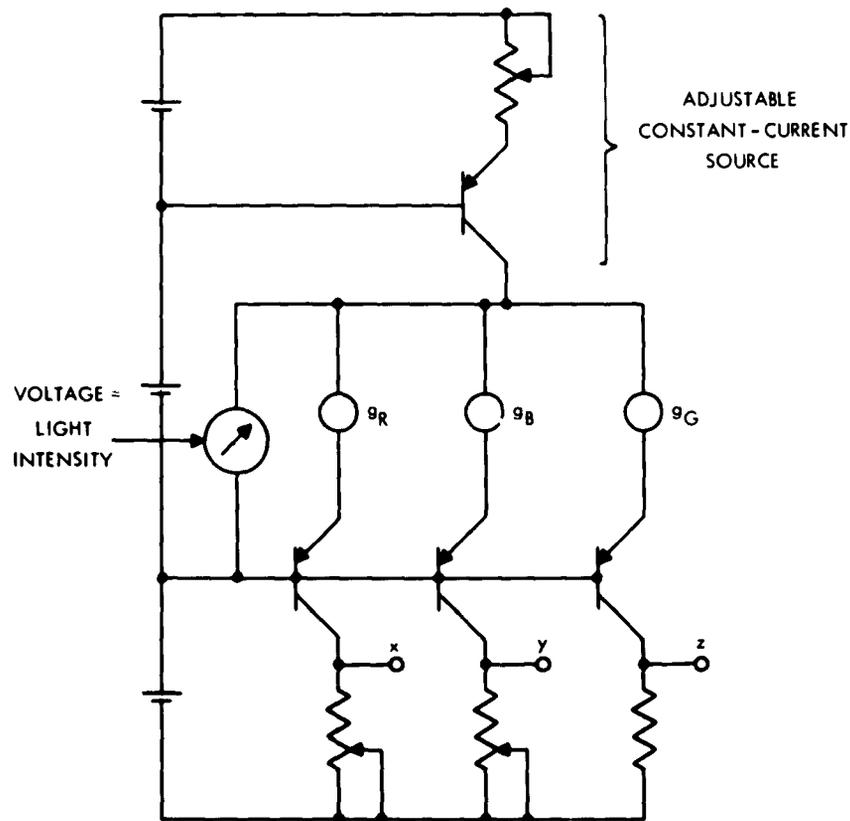

Fig. XXI-2. The Maggie A.

I should like now to show one of the variants of the device itself (see Fig. XXI-2). This version shows something that is like the Bezold-Brucke effect.

It seems to me that the general use of such quick and cheap computer analogues may be of use whenever one wants to construct a barycentric operation on a variety of data. So, for example, one might use strain gauges, thermistors, accelerometers, and all other such transducers that can be made to exhibit a conductance change that is a linear, or smooth and monotonic, function of the property to be sensed.

In general, when one is confronted with a spectrum of information, it is not a bad idea to look at that spectrum from three or more different views of the whole band. There are many advantages in so doing, as is clear from the comments on color vision. First, if you are comparing several different phenomena of different spectral composition, they have a certain distribution in the space relative to each other which tends to be preserved under translation, say, as with a continuously increasing bias between one end of the spectrum and the other. Second, phenomena that are undergoing a continuous change of spectrum can be distinguished sharply from those undergoing discontinuous changes.

Finally, it is occasionally of use to devise such a transformation on material which, at first glance, is thoroughly explored. Thus, for example, I see no reason why one cannot erect infrared color spaces and ultraviolet color spaces. But, more to the point, if, say, in cost accounting, or stock planning, one wishes to obtain a rough estimate and a variety of continuously variable factors determine the decision, the use of potentiometers and the circuit variant A (Fig. XXI-2), with as many potentiometers and transistors as are needed, would seem to be helpful.

One might feel that we have gone rather far afield from color vision. So we have. However, we now may ask whether such model-building is of help in thinking about physiology. I believe that it is, and I shall sketch a line of argument. Blue has a curious and anomalous position with respect to foveal color vision. It contributes little to the luminosity curve of the fovea; however, it has an enormous effect on perceived color. A mixture of yellow and blue that appears white remains white over a very considerable range of intensity If, however, at all levels of brightness you check the blue alone and the yellow alone for brightness, you find that they do not change in the same way. The yellow seems to get continuously brighter, whereas the blue does not seem to get brighter at nearly the same rate beyond a certain low level, but the mixture stays the same color. This suggests that the brightness function and color function are not rigidly joined, and certainly that the brightness encoding probably does not precede the encoding of color. For the various reasons suggested by Stiles, Rushton, and others, I believe two things about the receptors. The first is the fact that the bleach of a pigment molecule in a receptor causes the same increment of response in the receptor, no matter how many other molecules are bleaching at the time, that is, the receptor process is linear with light intensity over a considerable range. If it were not so, if a cone were logarithmic or followed a power law with respect to rate of light absorption, it is hard indeed to see what sort of compensatory mechanism would give us the barycentric operation that yields the trichromatic laws. I think that one has to be reasonable about this. Unless one can guess what sort of compensatory function could be used and how it is generated, one must not postulate an arbitrary sort of transformation and leave it to the good old cortex to untangle. The

second is the fact that many, if not all, of the cones have all of the pigments in them. Stiles and Rushton separately have very good arguments for this. What convinced me, however, was my experimentation with very small spots of light. Using a spot of less than 0.1-mm diameter at a distance greater than 1 meter (thus, less in visual angle than the size of a cone) and flashing the spot with a strobe light at a low rate, I was unable to detect differences for a just threshold red light over time. That is, it is very unlikely that the spot will hit the same cluster of cones (I am not supposing that the image of the spot was a cone in diameter — the dioptrics of the eye forbids this) twice in a row. Since the spot was kept in focus, I felt that it was unlikely to occupy much more than a twenty-cone area till it fell off markedly in intensity. At any rate, the spot did not change in character, that is, hue, intensity or brightness, from moment to moment. This, I felt, implied that all cones usually reported on the same thing. (Furthermore, the only two colors that I could distinguish were red and a sort of bluish green, but the red was always far more saturated. Yellow appeared white, and blue itself could not be seen until the intensity was increased so far that it was unlikely to occupy a limited area on the retina at threshold. For the same intensity a green spot could be detected more easily than a red spot, but the color could be more easily assigned to the red spot.)

Suppose that all cones bore all three pigments but in different ways. Let us set up an argument. A quantum of light is absorbed by a pigment molecule that is thereby bleached. This bleaching, however, is no transduction unless it is coupled somehow to a power amplifier. There are many possible kinds of amplification. For example, the bleaching may set up a propagating chemical change in the vicinity (as in the photographic process during development), and the resulting group of molecules is then released and diffuses to a place where, in turn, it can cause another change that leads eventually to a nerve firing somewhere. Or else the bleaching itself is an event that governs a membrane process so that an ionic current occurs, or an impedance changes, and so forth. I tend to believe, purely on faith, that the pigment molecules are adsorbed or in some way bound to the cone membrane, as some anatomists claim, and thus believe that the transduction is not mediated by several steps ending in a membrane change, but that the bleaching of a pigment molecule directly affects the membrane to which it is attached.

The way in which a continuous nerve membrane couples to itself from one patch to another is by virtue of the current flow between the patches, and the current is carried by ions. We may ask what sort of events are possible in the receptor end of the cone so that the cellular end and the pedicle are affected and can pass on information. If we believe that the information from the receptor is communicated electrotonically to the cell body and pedicle, and not by materials diffusing down the neck between cone and soma, then either current will pass from receptor to soma, or no event will be signalled farther on. Thus we enquire into what governs the flow of current through

membrane. There are two kinds of ions across a nerve membrane, those that are in equilibrium with the membrane potential and those that are not. (An ionic species is in equilibrium with a membrane potential when the potential between its activities on either side of the membrane equals the potential measured across the membrane.) If the resistance that a membrane offers to the passage of ions at equilibrium with the membrane potential (say K^+ or Cl^- in nerve fibers elsewhere) is lowered, then no current flows, and only the resistance to the flow of current changes. On the other hand, if the resistance to an ionic species not at equilibrium with the membrane potential is lowered, then it will tend to flow from the region of high activity on one side to the region of lower activity on the other side. This is an electric current through the membrane in one direction at the point where the resistance change has occurred, and diffusely back through the membrane in the other direction elsewhere. The membrane potentials of all other nerve elements which were measured are such that all ions of reasonable concentration and mobility (so that they are current carriers) are either at or close to equilibrium with it, or are far out of equilibrium with an activity potential in the opposite polarity to that of the membrane. I conceive that these events can occur at a transducer: either a nonequilibrium ion will be gated by an exciting event, an equilibrium ion or both. If only a nonequilibrium ion (say Na^+), is gated, current flows into the receptor and out through the rest of the membrane of the transducer and cell. If only an equilibrium ion (say K^+ or Cl^-) is gated, no current flows through the cell membrane. If both ionic channels are gated, the current that would ordinarily flow through the cell from the one is now attenuated by the shunting of the other. Thus we get "excitation" as current generation at a synapse or along a transducer membrane, and "inhibition" as a shunting, not itself capable of generating a signal but capable only of modifying excitations — essentially a divisive process. (I have commented elsewhere on the extension of this idea to dendritic trees.) However, in the cone, if we suppose that the bleaching of a green-absorbing molecule (Rushton's chlorolabe) is coupled to membrane in such a way that the bleaching step sets up a transient gating open of the Na^+ channel, then this pulse of current can be seen as a quantum of conductance change in the Na^+ channel so that it is linearly additive with similar quanta of g_{Na^+} from other sites in the same cone. Similarly, let us suppose that the red-absorbing pigment (Rushton's erythrolabe) is coupled in a similar way to K^+ or Cl^-, and so also is the blue-absorbing pigment (Rushton's cyanolabe). Do we not, then, have a good case (confected, it is true, out of no evidence) for which the device I have shown is a model? For in adjacent cones, only chlorolabe and erythrolabe need be reversed in their ionic-gating properties to generate the two necessary computations for the trichromatic theory. In another manner of speaking, in one cone, red light excites and both green and blue inhibit; in the other, green light excites and both red and blue inhibit. Thus blue can affect color independently of the way in which it affects luminosity, and its

effect on color would not depend on the subjective assessment of its brightness alone. (Willmer had something like this in mind.) You could probably make the followers of Hering very happy by having one cone in which red excited and green and blue inhibited, and another in which red and green both excited and only blue inhibited. The outputs would be what he requires for his "opponent" scheme. Any pair of different combiations work to generate the color plane, except those involving all pigments as all exciting or all inhibiting within the cone. Even blue can be made an occasional exciter without changing matters but giving it some luminosity function. The rules for combination by means of the nerve net can be made so explicit that only two adjacent cones have to be illuminated simultaneously for there to be no ambiguity between the colors. All other cones illuminated by the same light simply increase the resolution in distinguishing color, but do not change the color that is seen.

There are some distinct advantages in thinking that all cones bear all of the pigments. First, the dynamic range over which color can be held constant under change of brightness does not depend on carefulness in network design, and the operations are linear, as they should be. Second, the dynamic range of the receptor does not have to be as great and discriminable as it would have to be if each receptor had a separate pigment and displayed a linear output thereof. Third, it puts the color function before the luminosity function and thus is not dependent on it.

If broadband pigments are used, the vertical colors of the reference triangle must be imaginary. For, any pure spectral color will be represented by at least two numbers and a zero, while each of the vertices has one number and two zeros.

J. Y. Lettvin

Recommended Reading

Hermann L. F. von Helmholtz, Physiological Optics, Vol. 2 (Dover Publications, Inc., New York, reprint, 1962).

E. N. Willmer, Retinal Structure and Colour Vision (Cambridge University Press, London, 1946).

Committee on Colorimetry of The Optical Society of America, The Science of Color (Thomas Y. Crowell Company, New York, 1953).

James P. Southall, Introduction to Physiological Optics (Dover Publications, Inc., New York, reprint, 1961).

XXII. NEUROLOGY*

L. Stark	T. A. Rowe	S. F. Stanten
F. H. Baker	E. Sadler	A. Troelstra
R. W. Cornew	A. A. Sandberg	E. C. Van Horn, Jr.
H. T. Hermann	Susanne Shuman	G. L. Wickelgren
J. C. Houk, Jr.	J. I. Simpson	P. A. Willis
G. Masek	Gabriella W. Smith	S. Yasui
E. G. Merrill	I. Sobel	L. R. Young
R. Millecchia		B. L. Zuber

A. EYE CONVERGENCE

1. Improved Measuring Device

A pair of photocells (Clairex, CL-602), sensitive in the red region, have been mounted on eyeglass frames and connected into a bridge circuit. The subject's face is illuminated from above and at a distance of approximately 8 inches with a red light. These glasses offer several advantages over those previously used[1,2]; they are simple to construct; they are linear over a range of ±10 degrees; and, primarily, they reduce fatigue and strain caused by the rather intense visible light used with earlier glasses. The time constant of the improved glasses has been determined experimentally to be approximately 30 msec.

2. Results of Experiments

a. Step Response

Sample step responses appear in Fig. XXII-1. These are characterized by an
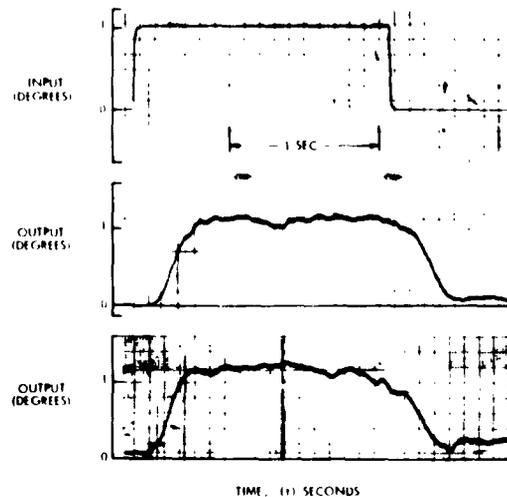
Fig. XXII-1. Step response.

average delay time of approximately 128 msec and an average rise time (between 90 per cent and 10 per cent amplitude) of 200 msec. The results were roughly the same, whether steps were predictable or unpredictable.

b. Frequency Response

Bode plots resulting from predfctable and unpredictable inputs are shown in Fig. XXII-2. These experiments were carried out by using a dc level corresponding to an apparent distance of 33.5 cm (angle of convergence, 5.1 degrees) and an ampli- tude corresponding to a change in apparent distance of approximately 5 cm (change of angle of convergence, 0.85 degrees). At this dc level, the apparent distance of the image is equal to the real distance. The dc level and amplitude were selected after
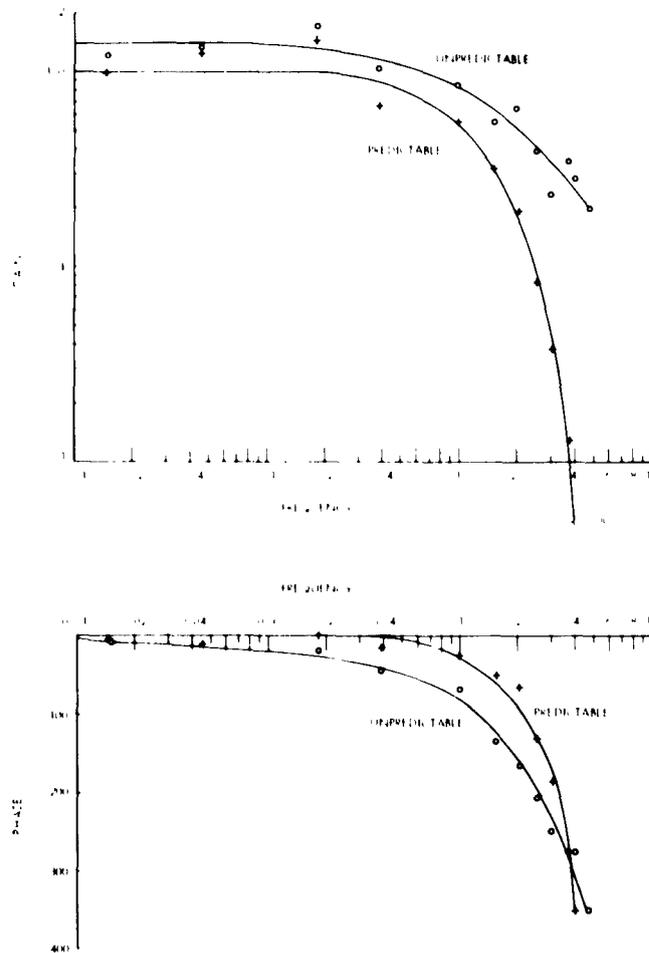
Fig. XXII-2. Bode plots.

it was found that using larger amplitudes, approximately 2.5 degrees, and dc levels differing from the real distance of the image resulted in the introduction of serious errors into the recorded data. When large amplitudes are used the subject is unable to keep the image in sharp focus over the entire range. In the part of the range in which the image appears out of focus, fusion is lost. The problem is further complicated if the subject attempts to adjust his accommodation for an image that might appear to be at a distance quite different from the real distance between the subject's eyes and the image. For these reasons a small amplitude about a dc level at which real and apparent distances are equal was selected to optimize accommodation.

The G. E. 225 computer was used to produce stimuli and to analyze responses. For frequencies less than 1 cps, results for the Bode plots were obtained in separate predictable and unpredictable input experiments. These gave characteristic gain and phase values for low frequencies. For frequencies greater than 1 cps, single frequencies and mixed-frequency groups were presented during the same experiment. In the latter case the order of presentation of single frequencies, groups, and frequencies within any given group was randomized.[3]

It is clear from the phase curves that a predictor apparatus operates to increase the phase margin near the -180-degree crossover frequency of the unpredictable curve.

c. High-Gain Oscillations

The unpredictable phase curve shows that the system has a phase lag of 180 degrees at a frequency of approximately 2.3 cps. If the loop gain of the system were sufficiently
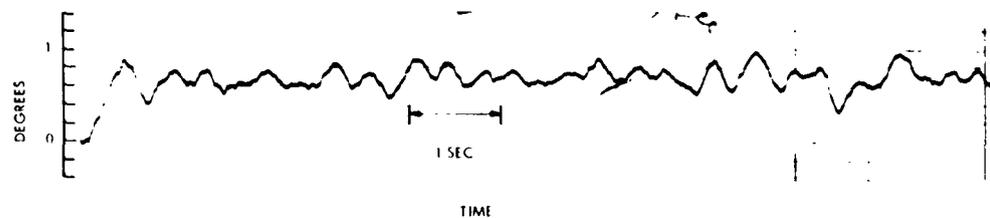
Fig. XXII-3. High-Gain oscillations.

high, one would expect the system to oscillate at this frequency. When an artificial loop was placed around the system and the gain increased, the system was found to oscillate spontaneously at an average frequency of 2.5 cps. Sample records are shown in Fig. XXII-3.

B. L. Zuber, A. Troelstra, L. Stark

References

1. L. Stark and A. Sandberg, A simple instrument for measuring eye movement, Quarterly Progress Report No. 62, Research Laboratory of Electronics, M. I. T., July 15, 1961, pp. 268-270.

2. G. P. Nelson, L. Stark, and L. R. Young, Phototube glasses for measuring eye movements, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M. I. T., October 15, 1962, pp. 214-215.

3. B. L. Zuber and L. Stark, Eye convergence, Quarterly Progress Report No. 68, Research Laboratory of Electronics, M. I. T., January 15, 1963, pp. 232-234.

B. ASSOCIATED EYE MOVEMENTS

The subject is presented with a target, as shown in Fig. XXII-4, by means of an X-Y recorder. The target movement, sinusoidal (0.3 cps) or stepwise (0.1 cps), is seen only with the right eye; the left eye is occluded by an opaque screen. The
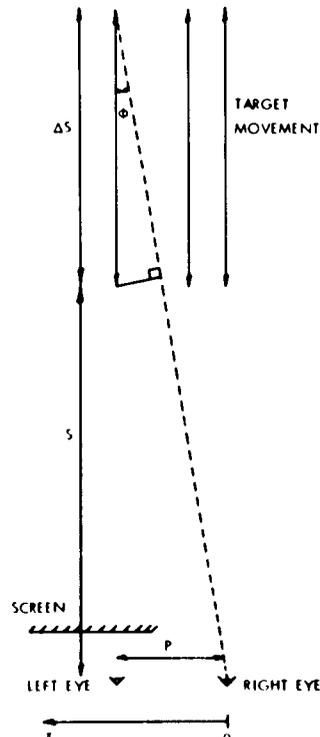
Fig. XXII-4. Experimental arrangement.

subject is asked to fixate the target and to keep it in focus. Movements of the left eye, which has no visual input, are measured. Visual input to the right eye may be resolved into a depth perception and a target displacement. Depth perception is a function of $\cos \phi$:

$$\text{Depth perception} = \Delta S \cos \phi = \frac{\Delta S(S+\Delta S)}{\sqrt{(S+\Delta S)^2 + X^2}} \approx \Delta S,$$

since $(S+\Delta S)^2 \gg X^2$ for the experimental conditions (S = 21 cm, ΔS = 15 cm, X_{max} = 6 cm). Consequently, the depth perception is more or less independent of the value of X.

Target displacement is a function of sin ϕ:

$$\text{Target displacement} = \Delta S \sin \phi = \frac{(\Delta S)X}{\sqrt{(S+\Delta S)^2 \pm X^2}} \approx \frac{(\Delta S)X}{S + \Delta S}.$$

Two types of associated movements of the left eye can result from the input to the right eye: accommodative convergence (AC) and associated tracking (AT).

1. Accommodative Convergence Movements

Accommodative convergence (or divergence) movements are defined as those movements resulting when the right eye changes its accommodation. For example, if the target is moved in the optical axis of the right eye, the left eye will follow roughly the target movement, even though the left eye receives no visual information and there are no movements of the right eye. With reference to Fig. XXII-4 and by assuming some gain for accommodative convergence, movement of the left eye which results from accommodative convergence will be

$$AC = G_{AC} \frac{(\Delta S)P}{(S+\Delta S)\sqrt{S^2 + X^2}} \frac{360}{2\pi} \text{ degrees.} \tag{1}$$

Gain will be frequency dependent, but in this experiment, in which low frequencies are used, that frequency dependence will be neglected. Furthermore, it will be assumed that G_{AC} depends only on the depth perception of the right eye and is independent of ϕ, and that G_{AC} is independent of additional visual information (for example, target displacement).

G_{AC} can be determined experimentally by measuring the movement of the covered and uncovered left eye when the target is moved in the optical axis of the right eye.

2. Associated Tracking Movements

Associated tracking movements of the left eye result when the right eye moves while tracking the target. These movements are opposite in sign compared with accommodative convergence movements. They are assumed to be proportional to the target displacement seen by the right eye.

$$AT = -G_{AT} \frac{(\Delta S)X}{(S+\Delta S) \sqrt{S^2 + X^2}} \frac{360}{2\pi} \text{ degrees.} \tag{2}$$

The definition of G_{AT} is similar to that for G_{AC}, except that G_{AT} is assumed to depend only on target displacement. G_{AT} is measured experimentally by presenting only a target displacement to the right eye and measuring the movement of the left eye, covered and uncovered. For three values of ϕ approximately 10 degrees apart the values of G_{AT} obtained are 1.25, 1.15, and 1.15.

3. Combined Associated Movements

In general, the right eye will receive information that consists of both depth perception and target displacement. If these two kinds of information contribute independently
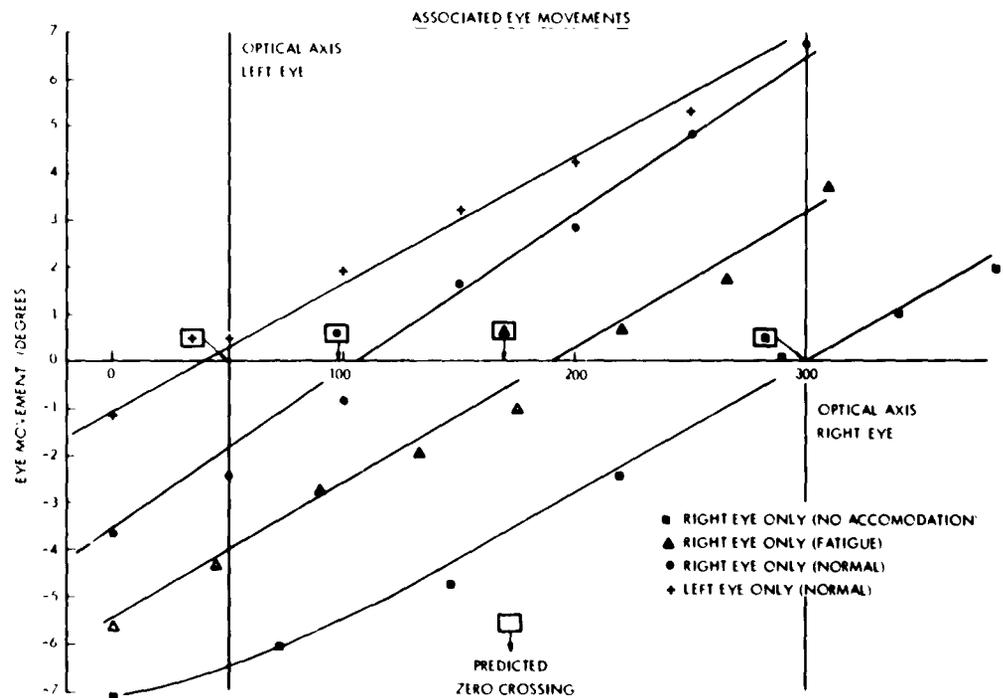
Fig. XXII-5. Associated movements of the left eye.

to the associated movement of the left eye, Eqs. 1 and 2 may be combined. In this experiment, then, the net movement of the left eye may be represented by

$$AC + AT = \frac{\Delta S}{(S+\Delta S) \sqrt{S^2 + X^2}} \frac{360}{2\pi} (PG_{AC} - XG_{AT}) \text{ degrees} \tag{3}$$

or

$$AC + AT = \frac{24}{\sqrt{440 + X^2}} (6G_{AC} - XG_{AT}) \text{ degrees.} \qquad (4)$$

From Eq. 4 it is clear that the value of X for which there will be no associated movement of the left eye is given by

$$6G_{AC} - XG_{AT} = 0. \qquad (5)$$

The value of $X_{zero\ crossing}$ as predicted by Eq. 5 was compared with the experimentally measured value for three different conditions of the eyes. The results are shown in Table XXII-1 and Fig. XXII-5.

Table XXII-1. Values of predicted $X_{zero\ crossing}$ for three different conditions of the eyes (left eye covered).

Condition of eyes	G_{AC}	G_{AT}	$X_{zero\ crossing}$ (cm)	$X_{zero\ crossing}$ (arbitrary abscissa of Fig. XXII-5)
Normal	0.95	1.17	4.9	97
Highly Fatigued	0.62	1.20	3.1	170
Not Accommodating	0	—	0	300

The experimental results clearly indicate that G_{AC} and G_{AT} are independent of each other, G_{AC} is greatly influenced by the degree of fatigue of the eyes, and G_{AT} is relatively free from fatigue.

A. Troelstra, L. Stark

C. ARTIFICIAL PHOTOSENSITIZATION OF THE CRAYFISH VENTRAL NERVE CORD

It has been known for a long time that muscle and nerve tissue after being stained with certain organic dyes becomes photosensitive. This report will consider the problem of classifying these artificial photic responses and associating these responses with physiological sites, presumably responsible for their origination.

1. Method

The artificial photoreceptor used in the experiment was the ventral nerve cord of the crayfish Cambarus astacus, which had been stained with Methylene blue (1:10,000)

for 90 minutes. The experiment was performed in an electrically insulated humidity chamber (100 per cent saturation, 10°C) which kept the nerve cord active for many hours.

The nerve cord was stimulated with a standard microscope illuminator, the light being focused to a spot 2.55 mm in diameter (0.06 lumen/mm^2).

The nerve pulses were recorded by using gross platinum hook electrodes. These pulses were fed into an electronic amplifying and windowing system and then displayed on a pen recorder.

2. Results of Experiment

Three characteristic features of the nerve responses were evident. The first was an excitatory response, an increase in the nerve pulse rate when the nerve was stimulated with light. The second was an inhibitory response, a decrease in the pulse rate when stimulated. The third was an "off" response, an increase in the pulse rate at the cessation of the stimulation.

A physiological location of the origin of each of these responses was made by carefully noting the site of stimulation associated with each response. An excitatory response occurred when the nerve cord was stimulated at a ganglion, while an inhibitory response, followed by an "off" response, occurred when a segment of axon was stimulated. Figure XXII-6 shows a recording of each of these responses and of a composite response obtained by stimulating a ganglion and a segment of axon simultaneously.

Using the electronic windowing technique, I found that only a small number of nerve fibers contributed to the photic response, namely the "B" fibers.

The minimum stimulus needed to evoke a response was found to be 0.4 second of
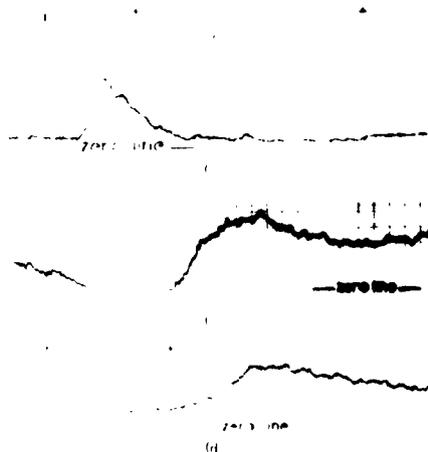
Fig. XXII-6. Response of artically photosensitized crayfish ventral nerve cord.
(a) Light stimulus (0.06 lumen/mm^2);
(b) Excitatory response of ganglion;
(c) Inhibitory and "off" response of axon;
(d) Composite response.

illumination from the microscope illuminator, which, on other terms, is approximately 25 photons/dye molecule (this is assumed to be the response of a single B fiber).

3. Minimum Photon Computation

The number of photons that are incident on the nerve can be calculated by using

$$\int_{\lambda_1}^{\lambda_2} N_I hc \, \frac{d\lambda}{\lambda^2} = E_T,$$

where

N_I is incident number of photons;

$\lambda_1 - \lambda_2 = 400 \, m\mu$-$700 \, m\mu$, wavelength range of visible spectrum;

$h = 6.62 \times 10^{-27}$ erg-sec, Planck's constant;

$c = 3 \times 10^{10}$ cm/sec, speed of light;

$E_T = kPAT$, total energy in ergs incident on nerve.

Here,

$P = 0.06$ lumens/min^2, power of light source;

$A = 2.6 \times 10^{-2}$ mm^2, stimulated area of nerve fiber;

$T = 0.4$ sec, minimum stimulation time;

$k = 1.61 \times 10^4$ (ergs/sec)/lumen, conversion factor.

Thus the incident number of photons is

$$N_I = 4.5 \times 10^{10} \text{ photons.}$$

The Methylene blue dye absorbs approximately 20 per cent of the incident illumination; therefore the number of photons absorbed by the dye is

$$N_A = 0.2 \, N_I = 9 \times 10^{10} \text{ photons.}$$

In a Methylene blue (1:10,000) solution the number of dye molecules per cubic centimeter is approximately

$$n = 1.8 \times 10^{17} \text{ molecules/cc.}$$

At this concentration and for a staining period of 90 minutes the nerve fiber takes up approximately 10 per cent of the dye. Therefore the number of dye molecules in the stimulated segment of nerve can be calculated from $N_M = 10\% \, nV$, where V, the volume of the stimulated nerve fiber, is 2×10^{-7} cc. Thus $N_M = 3.6 \times 10^9$ dye molecules.

We finally can calculate the minimum number of photons per dye molecule needed to evoke a response.

$$N_o = N_A/N_M = 25 \text{ photons/dye molecule.}$$

4. Discussion

The experimental results suggest the existence of two types of artificial photorecep-
tors, an excitatory one and an inhibitory one, the excitatory photoreceptor being local-
ized in the vicinity of the ganglia and the inhibitory photoreceptor being localized in the
vicinity of the axons. They also suggest that the receptors are associated primarily
with the B fibers that seem to have some generalized photosensitive properties.

R. Millecchia

D. HEAD MOVEMENTS

An investigation of the manner in which the rotations of the head aid human visual
tracking has been initiated. The purpose of the work to date has been threefold. First,
an efficient experimental procedure has been designed to minimize the time that is
necessary to record data from a particular subject. Second, a variety of excitations
has been applied to subjects in an attempt to characterize typical responses. Third,
under the hypothesis that allowing movement of the head serves to extend the frequency
response of the visual tracking system, a set of preliminary experiments with sums of
sinusoids used as excitations has been performed.

A narrow slit of light (7.5 X 0.25 cm) is projected by means of a slit projector onto
a Sanborn galvanometer-driven mirror, and is reflected onto a horizontal semicircular
projection screen. The projection screen has a radius of curvature of 2.7 meters and
the subject is seated at the center of screen curvature. Thus, a target angle is gener-
ated in proportion to the voltage that is denoted the target voltage. Figure XXII-7 shows
the relative positions of the major apparatus.
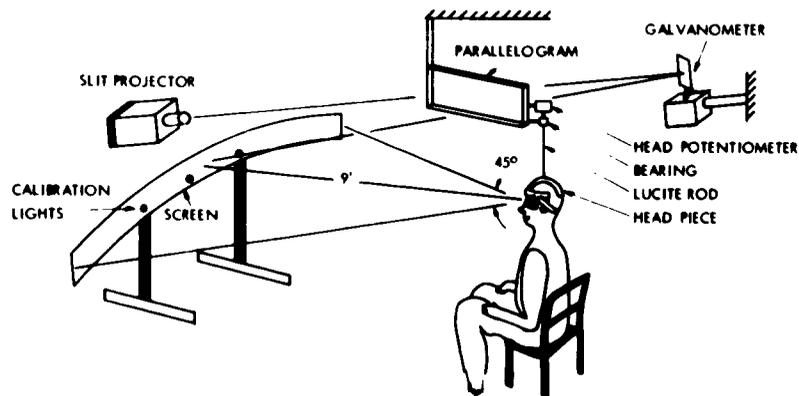
Fig. XXII-7. Relative positions of major apparatus.
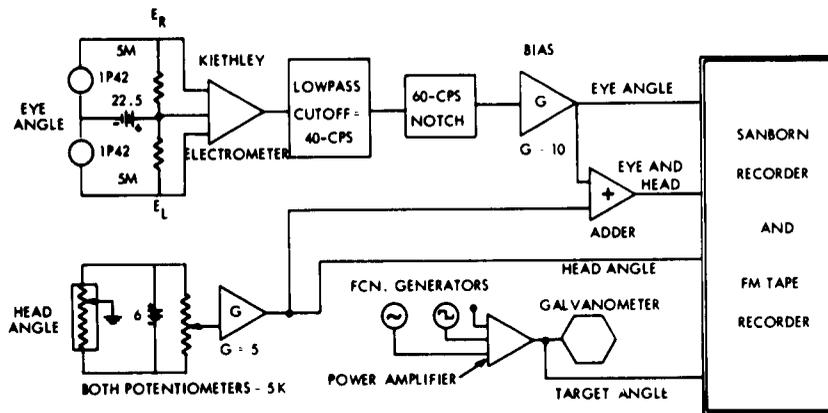
Fig. XXII-8. Simplified block diagram of eye angle and head angle instrumentation.

To measure head angle, a plastic headpiece was designed which snugly fits a variety of head sizes and shapes. The headpiece transmits information concerning head rotations in the horizontal plane to a potentiometer, which is connected in a bridge circuit. The headpiece also contains the two photocells and light bulbs used to measure lateral
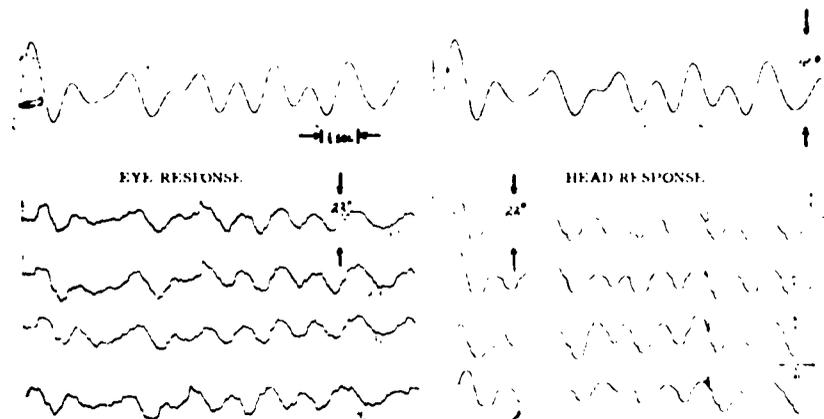
Fig. XXII-9. Four responses to sum of 5 sinusoids.

eye angle with respect to the head.[1] The method of constraining the head with the headpiece does not interfere at all mechanically, nor to a great extent psychologically, with the natural movements of the head. Figure XXII-8 shows a simplified flow diagram of the head- and eye-angle measurement instrumentation.

By prerecording the various excitations, sums of sinusoids, on FM tape and by recording the three responses (head, eye, and head-plus-eye angles) simultaneously

on tape, an existing G. E. 225 computer program (MITMR) was used to analyze sequentially the head, eye, and head-plus-eye angles. Figure XXII-9 shows four head-and-eye angle responses to a sum of 5 sinusoids.

1. Single Sinusoid Responses

Several characteristics occur consistently in the single-sinusoid responses. Qualitatively, for a constant-amplitude target angle the individual gains of eye and head angle vary with time; however, the head-plus-eye angle gain remains constant with time. Zero shifts in eye angle and head angle are as large as 20 per cent of the respective amplitudes, but are complementary, so that no zero shift occurs in the head-plus-eye response. Head angle gain (or normalized amplitude) varies with target amplitude; furthermore, the Bode asymptotic intersection frequency decreases as target-angle amplitude decreases.

2. Square Step Responses

Figure XXII-10 shows a typical response to a square step of target angle. Characteristic discrete jumps in head angle occur with intervening periods ranging from
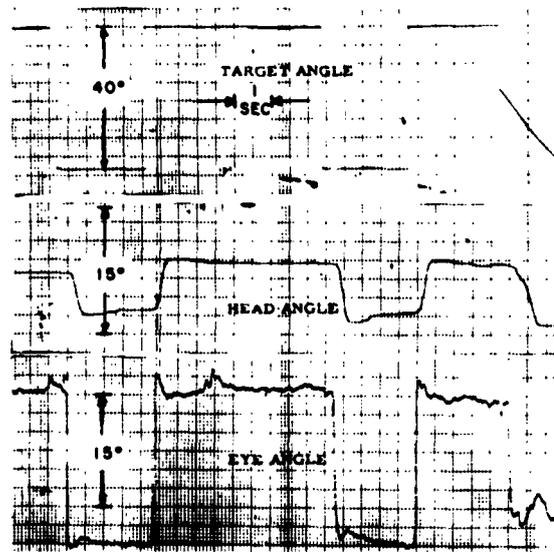
Fig. XXII-10. Typical responses to square step input.

200 msec to 600 msec and amplitudes of 0.1-0.4 degrees.[2] Latencies of eye response are not dependent upon whether the head is free or fixed, but range from 200 msec to 300 msec. Average response of head movement is 400 msec.

G. Masek

References

1. L. Stark and A. Sandberg, A simple instrument for measuring eye movements, Quarterly Progress Report No. 62, Research Laboratory of Electronics, M. I. T., July 15, 1961, pp. 268-270.

2. L. R. Young, A Sample Data Model for Eye Tracking Movements, Sc. D. Thesis, Department of Aeronautics and Astronautics, M. I. T., May 11, 1962.

E. MODIFICATION OF INPUT PARAMETERS FOR EKG PATTERN-RECOGNITION ANALYSIS

The G. E. 225 computer with integral analog-to-digital and digital-to-analog equipment has been incorporated into an experimental system (Fig. XXII-11) for remote on-line medical diagnostic study. This system applies pattern-recognition techniques to the classification and automatic diagnosis of clinical electrocardiograms.

A program for the IBM 7090 computer has been developed for pattern recognition which uses a system of multiple adaptive matched filters. The recognition system appears to operate in a manner that is similar to the human pattern-recognition process that is the conventional means used for EKG analysis. The filter analyses are made in the time domain that is such that no Fourier analysis is required. The over-all program involves a variety of normalization, weighting, comparison, decision, modification, and adapting operations.

A complete diagnosis of the EKG requires an analysis of all three vector components $(X(t), Y(t), Z(t))$ of the spatial vector signal. Recent investigations of the pattern-recognition system treat only one vector component (X, Y or Z) of the signal. Methods have been investigated to treat the vector components as related patterns and to associate pertinent information from one component to another to form one function (X, Y, Z). By a reduction of input data to a form that is more suitable for the pattern-recognition system, the computer diagnosis of the EKG can be simplified.

The function (X, Y, Z) must meet the following criteria to be useful in the pattern-recognition system:

(a) It must be similar for each set of data corresponding to one particular diagnosis.

(b) It must be unique for data corresponding to different diagnoses.

The G. E. 225 computer has been programmed to determine a variety of geometrical parameters of the EKG spatial curve. The functions being investigated include spherical coordinates, first derivative of arc length, and curvature. Each parameter is a function of time so as to coincide with the recognition system. Figures XXII-12 and XXII-13 show typical results from two different data sets.
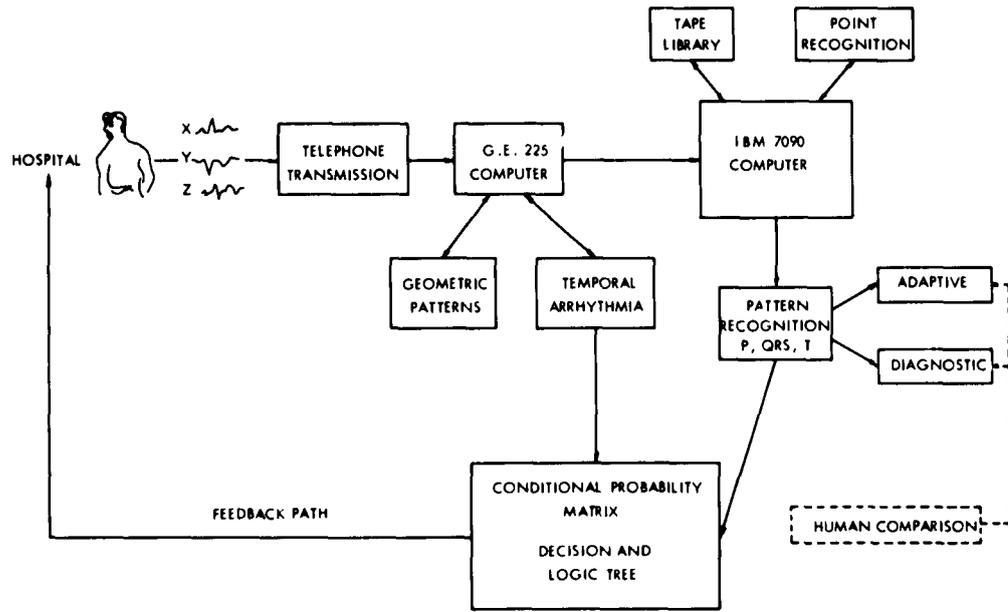
Fig. XXII-11. The G. E. 225 computer with integral analog-to-digital and digital-to-analog equipment. (Geometric patterns modify input parameters.)
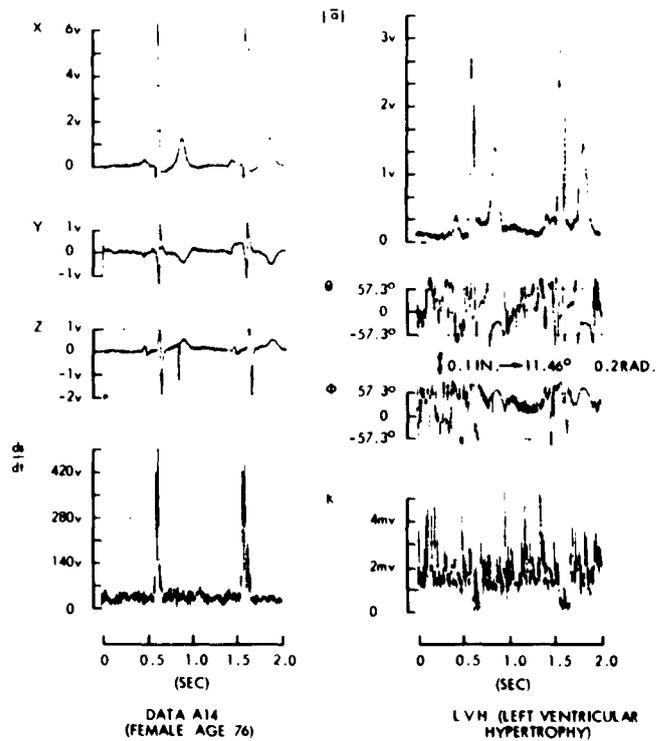
Fig. XXII-12. The EKG vector and associated functions for a female of age 76 years with left ventricular hypertrophy.
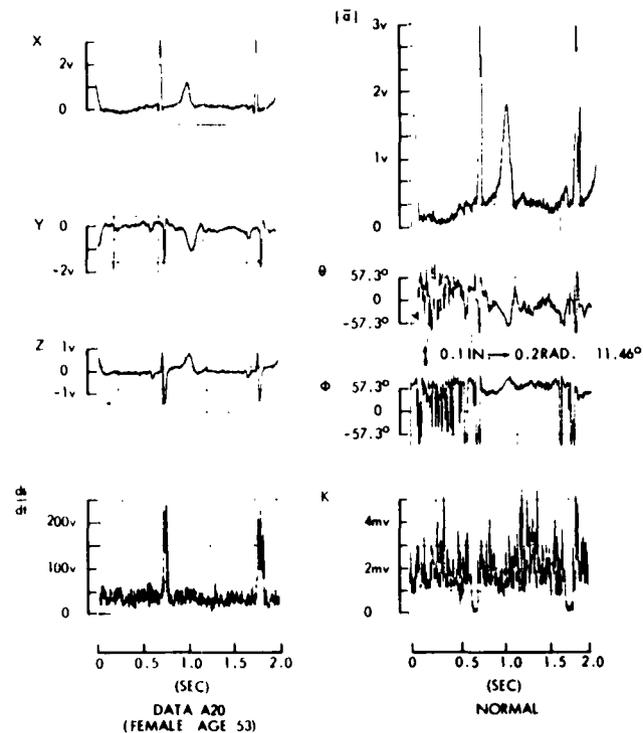
Fig. XXII-13. The EKG vector and associated functions for a normal
female of age 53 years.

The spherical coordinates (spatial magnitude $|\bar{a}|$, angle of azimuth θ, and angle of elevation ϕ), the first derivative of arc speed $\frac{ds}{dt}$, and the curvature function K are all potentially valuable functions. However, further investigations of computation methods preserving signal-to-noise ratios are required. The aim is to transform the data so that we obtain three redundant vectors or alternatively that all the signal is in one vector, leaving only noise in the others. In this way the pattern-recognition program can operate only upon any one of the three redundant vectors or on the signal-bearing single vector, respectively.

We are now collecting many examples of these functions so that they may be tested by using the pattern-recognition program in order to obtain a meaningful interpretation of the data-reduction techniques.

E. Sadler, L. Stark, J. Dickson,
I. Sobel, G. Whipple

F. SOME EXPERIMENTAL PROPERTIES OF PUPILLARY NOISE

It was noticed during the investigation of pupillary noise that there is an increase in the rms noise level as the illumination level is increased. In order to investigate
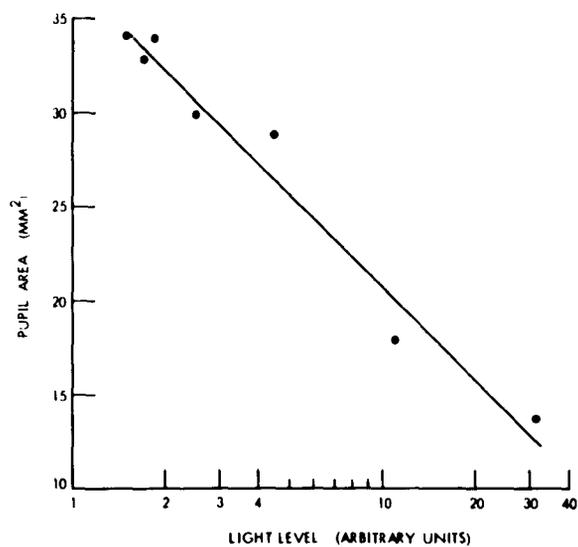
Fig. XXII-14. Pupil average area $\overline{A}$ vs light level L.
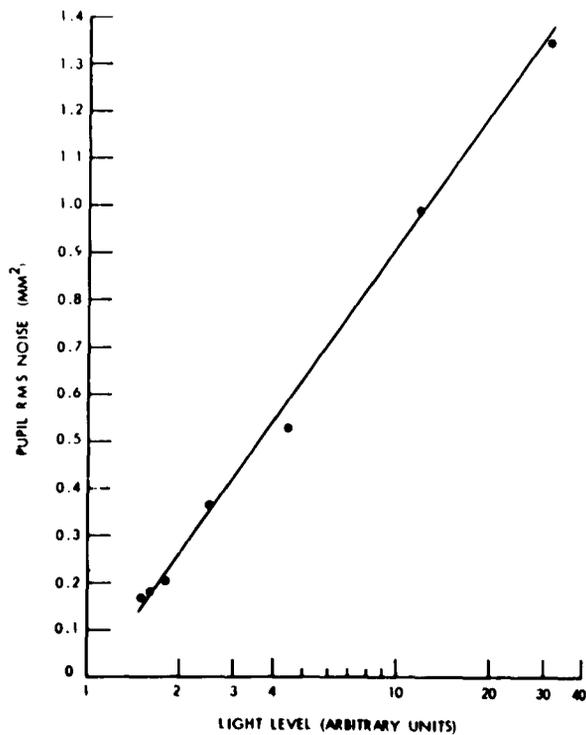
Fig. XXII-15. Pupil rms noise σ vs light level L.
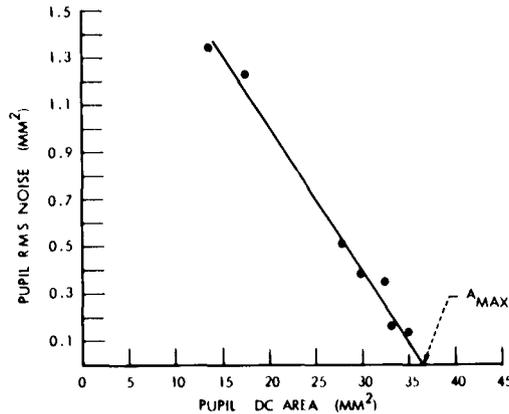
Fig. XXII-16. Pupil rms noise σ vs pupil average area $\bar{A}$.

this effect more fully, a program was written for the G. E. 225 computer which computed the rms and average of the pupil area signal. The experiment was performed by presenting the subject a constant light level and by measuring the rms and average area for approximately 60-80 seconds. A sampling rate of 50 samples per second was used.

Figure XXII-14 shows the average area, $\bar{A}$, as a function of light, L, on a semilog scale. The logarithmic dependence of $\bar{A}$ upon L was quite expected because of various experiments on the pupil which show a logarithmic dependence of the pulse height response as a function of input pulse height. The unexpected result was the logarithmic dependence of the rms noise upon light level shown in Fig. XXII-15.

Figure XXII-16 is a graph of rms noise vs $\bar{A}$ on a linear scale. Notice also the almost perfect linear relation between rms noise and $A_{max} - \bar{A}$. A_{max} is the maximum attainable average pupil area. This result might indicate that the noise is of a multiplicative nature.

In order to reinforce this hypothesis a second experiment was performed in which the pupil was stimulated by a 2-cps sinusoid and the autocorrelation of the pupil area response studied. If the hypothesis that

$$A_{max} - A = C_1 n(t) I + C_2 I \tag{1}$$

is true, where C_1 and C_2 are constants and I is some monotonic function of L, possibly log L, and if we assume that $L \approx B + D \sin \omega t$, then the autocorrelation of area signal, $R_A(\tau)$, after elimination of the constant A_{max} term, should be

$$R_A(\tau) = \frac{C_1^2 D^2}{2} R_n(\tau) \cos \omega\tau + \frac{C_2^2 D^2}{2} \cos \omega\tau + C_1^2 B^2 R_n(\tau) + C_2^2 B^2, \tag{2}$$

where $R_n(\tau)$ is the noise source autocorrelation function.

If the additive term $(C_1^2 D^2/2) \cos \omega\tau$ and constant term $C_2^2 B^2$ are subtracted from $R_A(\tau)$, we should obtain

$$R_A(\tau) - \frac{C_1^2 D^2}{2} \cos \omega\tau - C_2^2 B^2 = R_n(\tau) \frac{C_2^2 D^2}{2} \cos \omega\tau + B^2 D^2 R_n(\tau). \tag{3}$$

If $R_n(\tau)$ is assumed to be approximately exponentially decreasing, then Eq. 3 is the sum
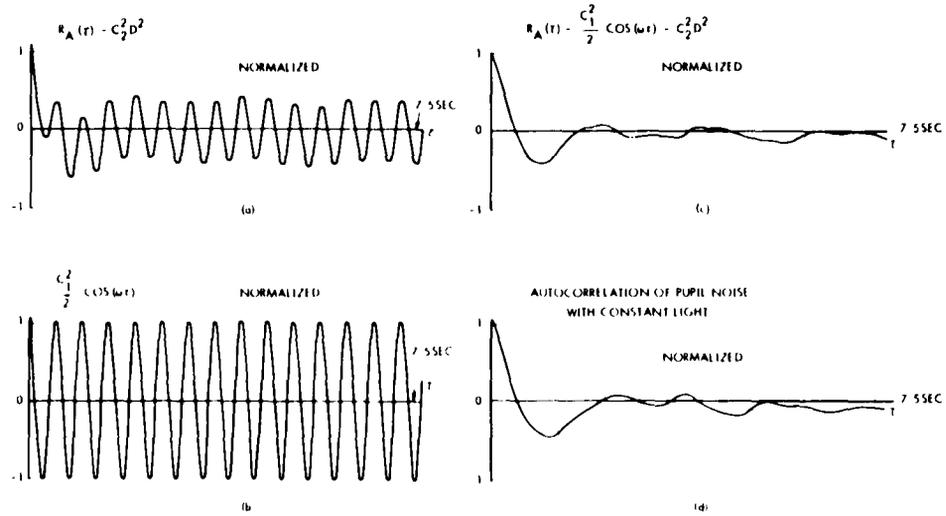
Fig. XXII-17. Autocorrelation functions of (a) pupil signal plus noise; (b) pure signal; (c) noise plus signal minus signal; and (d) noise alone.

of two terms. The first represents a signal-noise interaction term and the second, a pure noise term. The signal-noise interaction term should be an exponentially damped cosine.

A program was written for the G. E. 225 computer which accepted the autocorrelation of the pupil area, as well as the autocorrelation of the driving sinusoid, for a frequency reference, and

$$R_A(\tau) - \frac{C_1^2 D^2}{2} \cos \omega\tau - C_2^2 D^2$$

was computed. This program calculated $C_1^2/2$ by measuring the average amplitude of oscillation of the last half section of the autocorrelation function. The experimental
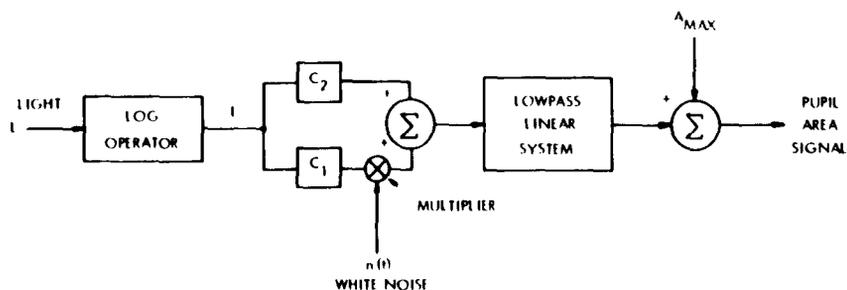
Fig. XXII-18. Simplified model to explain results of pupil noise experiments.

results are shown in Fig. XXII-17. Each autocorrelation was made up of 4 pieces of datum, each 60 seconds in length and bandpass filtered (0.03-10 cps) in order to eliminate low-frequency trends.

It is important to note that in Fig. XXII-17c we do not get a damped oscillatory behavior as might be predicted from Eq. 3. As a matter of fact, when the additive portion of the signal was subtracted, all that remained was an autocorrelation that looked like the autocorrelation of pupillary noise under constant-illumination conditions, as shown in Fig. XXII-17d for comparison purposes. This result might be described by the model shown in Fig. XXII-18 in which a lowpass filter is added after the noise is introduced into the system, in order to obtain the observed spectral characteristics.

S. F. Stanten, L. Stark

G. OPTOKINETIC NYSTAGMUS: DOUBLE STRIPE EXPERIMENT

This work is a continuation of previous studies of the optokinetic nystagmus in man.[1,2] Earlier reports contain descriptions of both the apparatus and
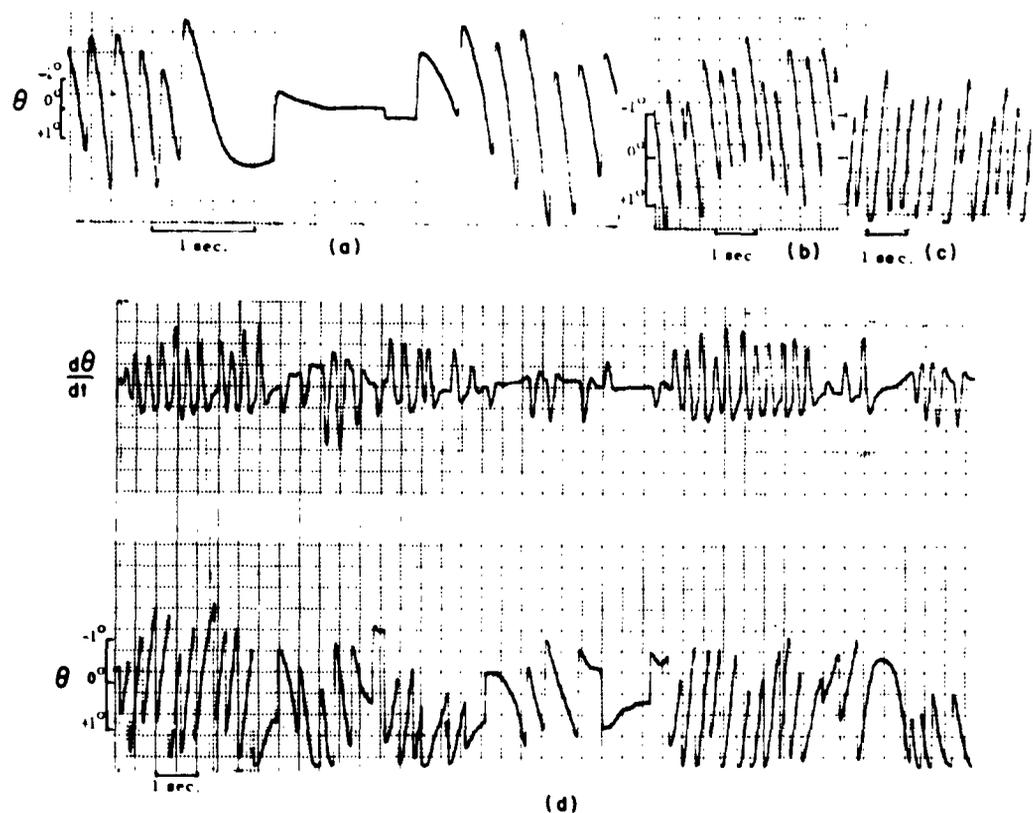
Fig. XXII-19. Records of optokinetic nystagmus.

the optokinetic nystagmus reflex.[1-4]

In the "unfixated" case (there is no fixation point) the subject is directed only to maintain a forward gaze. His visual field is dominated by a pattern of moving, vertical stripes. With reasonably high stripe velocities (w_f = 40-50°/sec), the nystagmus is occasionally halted, the last pursuit movement smoothly decelerating. After a period of slow or zero velocity, the first one or two pursuit movements generally accelerate smoothly, later pursuit movements in the optokinetic nystagmus sequence having constant velocities. The deceleration, acceleration, and segments of slow or zero eye movement interrupting the optokinetic nystagmus can be seen in Fig. XXII-19a.

Figure XXII-20a shows the same data as Fig. XXII-19a. The saccadic returns have been cut away, and the smooth pursuit movements pieced together and thus the independence of pursuit and saccadic segments of optokinetic nystagmus is emphasized. The small peak on the horizontal section between the sections of optokinetic nystagmus shows that the independence is not complete. In Fig. XXII-19b and 19c two additional records of nystagmus are shown. Figure XXII-19b is the result of stripes moving from left to right, while Fig. XXII-19c results from left-moving stripes, as indicated by the different directions of the slow phases.

(a)   (b)
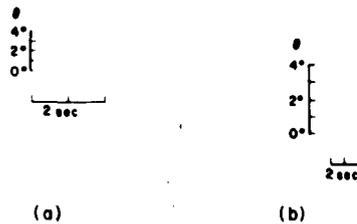
Fig. XXII-20. Constructed sequences of slow phases.

Figure XXII-19d shows eye movements resulting from the two sets of stripes presented simultaneously that generated the optokinetic nystagmus in Fig. XXII-19b and 19c. In this experiment, the stripe velocities were both approximately 9°/sec. The filtered derivative is shown at the top of Fig. XXII-19. Discrete optokinetic nystagmus segments in both directions are connected with accelerating and decelerating pursuit movements that are similar to those noted in the single-stripe experiments. Occasionally, the decelerating-accelerating movements are connected without an intervening saccade, resulting in a hump between the two directions of optokinetic nystagmus.

Figure XXII-20b shows the data in Fig. XXII-19d with the saccades taken out and the

smooth movements joined. The changes in optokinetic nystagmus that result from left-moving stripes becoming right-moving stripes can be seen as a smooth change in velocity from positive to negative.

E. G. Merrill, L. Stark

References

1. L. Stark and G. Nelson, Optokinetic nystagmus, Quarterly Progress Report No. 64, Research Laboratory of Electronics, M. I. T., January 15, 1962, pp. 326-328.

2. G. P. Nelson and L. Stark, Optokinetic nystagmus in man, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M. I. T., July 15, 1962, pp. 366-369.

3. L. Stark and A. Sandberg, A simple instrument for measuring eye movements, Quarterly Progress Report No. 62, Research Laboratory of Electronics, M. I. T., July 15, 1961, pp. 268-270.

4. G. P. Nelson, L. Stark, and L. R. Young, Phototube glasses for measuring eye movements, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M. I. T., October 15, 1962, pp. 214-216.

XXIII. SENSORY AIDS RESEARCH[*]

Prof. S. J. Mason J. Dupress J. T. Lynch
Prof. D. E. Troxel J. W. Holcroft M. A. Pilla
W. L. Black J. H. Hufford L. R. Shulman
G. Cheadle W. G. Kellner J. M. Topaz
R. C. Levine

A. SENSORY READING AID FOR THE BLIND

1. Introduction

In pursuit of the long-range objective of providing better sensory reading aids for the blind, research has been carried out in the field of the tactual and kinesthetic information-transfer process by using a stenotype machine operated in reverse. A system has been designed and built in which punched paper tape, coded from stenotype output tapes, feeds a decoder, which in turn provides signals for a key-actuator console to move the keys of a stenotype machine. A stenotype-trained subject has been taught to "read" the output (key movements under the fingers), as if she were operating the machine in its normal mode, and to respond by saying the words presented. Advantage has been taken of the fact that the subject was well versed in the stenotype code (a sort of redundance-reducing phonetic shorthand using ordinary typed letters) by making the output key movements correspond to the same code, with minor restrictions.

There is need for a faster sensory reading aid for the blind. At present, the only such system in general use is Braille. This system is "read" by scanning tactually with the index finger a succession of spatial cells, each of which contains a configuration of raised bumps (or absence of same) in six positions in the cell. In the elementary Braille system, called Braille I, each configuration, generally speaking, corresponds to a single letter of the alphabet. Since this obviously limits reading to a letter-by-letter flow, it is undesirably slow. A faster system, which is more difficult to master, is called Braille II, which uses abbreviations and encodes some whole words and common letter groupings into single cells. Even Braille II is not acceptable as the final answer for a sensory aid, however, because of its slow speed compared with visual reading. Normal prose can be read visually at a rate of approximately 300-400 words per minute, while the rate for Braille II is only approximately 70-90 words per minute.

The stenotype machine, operated in reverse, was chosen as a kinesthetic/tactile stimulator for three reasons: First, sensory communication techniques using movement of and pressure on the fingers appeared to hold greatest promise. Second, the use of many fingers simultaneously as receptors appeared in a general way to increase

the sensory channel capacity greatly as compared with single-finger stimulation (such as Braille). Third, the stenotype system reduces the redundancy of English prose into a phonetic code with some smoothing of the information rate, which was thought to lend itself to the achievement of speed in sensory reading.

Fig. XXIII-1. Example of stenotype machine output.

The stenotype machine is a device that looks like a small typewriter with an unusual keyboard, whose keys cause a Roman type to print on an adding-machinelike paper roll (See Figs. XXIII-1 and XXIII-2). Trained stenotypists transcribe from voice to paper by depressing one or more keys simultaneously, causing the printing of letters on the paper in an almost phonetic code at approximately an average of one word per line (a new line feeds at every stroke of the keys). In order to be considered a qualified stenotypist for normal secretarial-stenographic work, an operator must be able to transcribe at least 125 words per minute, and in order to be qualified as a court stenographer an operator must be able to transcribe at least 200 words per minute.
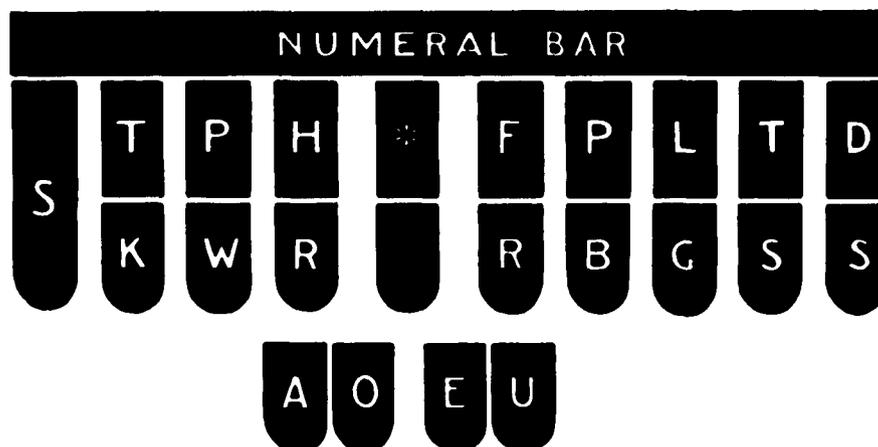
Fig. XXIII-2. Stenotype machine keyboard.

Experts can transcribe 300, or sometimes more; contests involve speeds over 350.

2. Experiment

In this research project the stenotype machine was operated in reverse. That is, a message was properly encoded on punched tape, which was fed into an electronic decoder and keyboard actuator in such a manner that the keys of the machine moved up and down as if the succession of words constituting the message were being transcribed by a disembodied operator. The subject, a trained stenotypist, kept her fingers in the proper "rest position" on the keyboard, noted the key actuation, and responded by saying the words that would have been transcribed had she been operating the machine herself.

When in the rest position the stenotype operator's fingers are in contact with 19 of the 23 keys of the machine. In this experiment, test words were chosen which used only the 19 characters that could be sensed without moving the fingers. Very little generality was lost in that the remaining four keys represented little-used characters.

Being phonetic in nature, the stenotype code uses generally one line per syllable, exclusive of abbreviations (of which there are many). Because individual syllables of multisyllable words can in many cases be construed (phonetically) to be words in themselves, multiline words added a confusion factor that was undesirable during the learning process. Accordingly, words were chosen for the speed test which could be encoded into only one line.

The subject was given a familiarization period of approximately 18 hours with punched tapes coded from three different sources: a list of the 1000 most-used English words, a jury trial transcript, and three reading-comprehension stories. During this

period three modes of operation were tried: monitor-operated foot pedal, subject-operated foot pedal, and continuous operation with adjustable repetition rate. After a few hours of very basic familiarization the one-shot modes with the foot pedal became less satisfactory than the continuous because the subject tended to search for all depressed keys rather than to sense the over-all pattern as one "chunk" of information. The familiarization tapes also had what turned out to be undesirable features of multi-syllable words and the inclusion of non-sensed keys. At the initial slow reading rates, context proved to be of no help. For these reasons new tapes were made to be used in speed trials.

It was felt that it would be impossible to obtain definitive quantitative results in terms of precise information at rates from an experiment that was built around English prose encoded in a still unanalyzed code. Nevertheless, it would be desirable to obtain results that would relate in some manner (however subjectively) to the real world. Therefore it was decided to make three context-free lists of English words of varying stenotype difficulty and to see how fast each list could be "read" by the subject.

3. Discussion of Results

Three sets of one-line English words that used only the proper 19 keys were chosen at random from the previous tapes. The "easy" list contained 175 words for the most part of no more than four stenotype characters each. The "intermediate" list contained 175 words of (generally) from four to six stenotype characters each. The "hard" list contained 130 words of four or more (up to ten) stenotype characters each. Tapes were made from these lists and 12 hours of familiarization practice performed.

In reading these lists the subject made a number of different errors. These ranged from the simple omission of the beginning or final sound of a word to the loss of the entire word by missing a crucial stenotype character of a highly encoded con-sonant sound. These errors obviously have much different significances, so that it would be difficult to give meaning to an "error-rate." Accordingly, it was decided to operate essentially error-free. "Essentially" means that no more than two of the first kind of error described above would be tolerated in a given list.

To eliminate the possibility that context was used in the reading, scrambled ver-sions of the lists were made. A new scrambled version was used for each of the final speed tests. Using the above-mentioned error criterion, the subject achieved the following speeds:

"Hard" List	31 words per minute
"Intermediate" List	40 words per minute
"Easy" List	62 words per minute

The speeds achieved suggest that the technique of multifinger stimulation with a good redundancy-reducing code for the reading of English has considerable promise

and may be a means whereby speed comparable to sight reading can be attained.

G. Cheadle

B. AN UPPER BOUND TO THE "CHANNEL CAPACITY" OF A TACTILE COMMUNICATION SYSTEM

1. Communication System

The system reported on here consists of a device that can elicit a pressure response from the finger tip. The stimuli can be applied to any combination of 6 fingers by 6 poke probes, which are small rivets that move approximately 3/8 inch. The subject can feel the pressure of each of the 6 probes, thereby obtaining 6 bits of information per stimulus. Figure XXIII-3 shows the arrangement of the probes relative to the fingers, and illustrates the way in which a stimulus was interpreted as a two-digit octal number.

Fig. XXIII-3. Probe arrangement.

If the subject could receive a single-probe stimulus 6 times a second, then the information rate would be 6 bits/sec, provided that he made no errors. Likewise, if he received a 6-probe stimulus once a second, the information rate would be 6 bits/sec. From a communication viewpoint, both alternatives are equally good. The purpose of our experiment was to determine what ensemble size provides the greater mutual information rate.

The stimuli were presented tachistoscopically, to allow as long a time as necessary for the subject to decode the stimuli. The skin has a finite relaxation time, however, and there is an "afterimage" effect. One can easily observe this phenomenon by poking his finger tip with a pencil point and noting the response after the pencil has been removed.

To solve this problem, a mask consisting of all 6 poke probes was applied after each

and may be a means whereby speed comparable to sight reading can be attained.

G. Cheadle

B. AN UPPER BOUND TO THE "CHANNEL CAPACITY" OF A TACTILE COMMUNICATION SYSTEM

1. Communication System

The system reported on here consists of a device that can elicit a pressure response from the finger tip. The stimuli can be applied to any combination of 6 fingers by 6 poke probes, which are small rivets that move approximately 3/8 inch. The subject can feel the pressure of each of the 6 probes, thereby obtaining 6 bits of information per stimulus. Figure XXIII-3 shows the arrangement of the probes relative to the fingers, and illustrates the way in which a stimulus was interpreted as a two-digit octal number.

LEFT HAND   RIGHT HAND

Fig. XXIII-3.   Probe arrangement.

If the subject could receive a single-probe stimulus 6 times a second, then the information rate would be 6 bits/sec, provided that he made no errors. Likewise, if he received a 6-probe stimulus once a second, the information rate would be 6 bits/sec. From a communication viewpoint, both alternatives are equally good. The purpose of our experiment was to determine what ensemble size provides the greater mutual information rate.

The stimuli were presented tachistoscopically, to allow as long a time as necessary for the subject to decode the stimuli. The skin has a finite relaxation time, however, and there is an "afterimage" effect. One can easily observe this phenomenon by poking his finger tip with a pencil point and noting the response after the pencil has been removed.

To solve this problem, a mask consisting of all 6 poke probes was applied after each

stimulus pulse. The purpose of the mask is to "erase" the afterimage effects of the skin. The subject is again given unlimited time to decode the stimulus. The time per stimulus is defined as the time from the start of the coded symbol to the start of the mask.

2. Preliminary Tests

To test the effectiveness of the mask a binary tape that did not contain a mask was prepared. When an ensemble size of 6 was used the error rate in tests with the mask was twice the error rate in tests without the mask. This result indicated that the subject did indeed receive some information after the probes had retracted, thereby warranting the use of the mask. The error rate, rather than the mutual-information rate, was used because of the difficulty in defining the duration of the stimulus in the maskless experiment. In the rest of the experiment, however, the stimulus duration was well defined, and comparisons were made by using the mutual information rate.
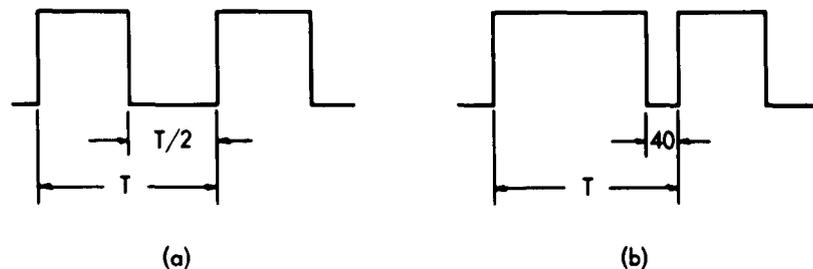
Fig. XXIII-4. Timing of stimuli.

In the early part of the experiments the time that the actual symbol was touching the finger tip was one half of the defined stimulus duration (see Fig. XXIII-4a). It was discovered later that much higher rates could be achieved by making the interval between the symbol and the mask as short as possible (40 msec) while keeping the defined stimulus duration constant (see Fig. XXIII-4b). The resultant increase in the information rate is shown in Fig. XXIII-5. Note that the information rate does not change so radically for the ensemble size of 4, and not at all for the ensemble size of 2. The reason for this is that the defined time of the stimulus duration was approximately 100 msec for the ensemble size of 2. In the last case parts (a) and (b) of Fig. XXIII-4 would be too much alike for a significant change to be noted; however, the time for the ensemble size of 6 was approximately 400 msec, and thus the top two parts of the figure would be decisively different.
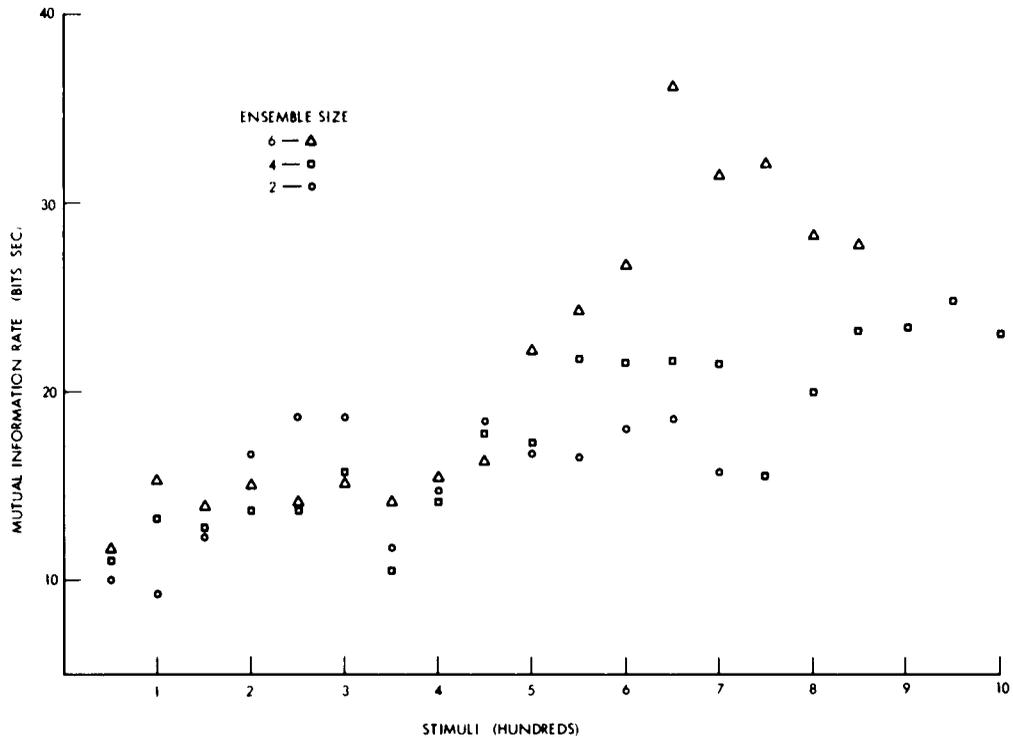
Fig. XXIII-5. Learning curves.

3. Results and Conclusions

The information rate was estimated for sequences of 50 random two-digit octal numbers. We felt that 10 per cent error would be reasonable for such a device; therefore the duration of the stimulus was adjusted to allow the error to approach 10 per cent. Learning curves for each of the three ensemble rises are plotted in Fig. XXIII-5.

To determine the asymptote of each of the learning curves, an average value after leveling off was calculated. The mutual information is not a linear function of the transition probaiblities; therefore the average was evaluated by constructing a channel matrix from the last 200 stimuli. The asymptote average for each of the ensemble sizes is given below.

Ensemble Size	I(xy) Bits/sec
6	29.5
4	23.2
2	16.2

From the experimental data one must conclude that the ensemble size of 6 gives a significantly higher mutual-information rate than either the ensemble size of 4 or 2. Although the margin between ensembles sizes of 4 and 2 is not as large, we conclude that for such small ensemble sizes the mutual-information rate increases with an increase in the ensemble size.

It must be noted that in a tachisotoscopic experiment such as this the calculated information rate is an upper bound on the actual information rate, and that this upper bound should not be compared with any reading rate. The upper bound, however, does provide a basis for comparing various parameters of the channel.

Further details can be found in the author's thesis.

J. T. Lynch

References

1. J. T. Lynch, Variation of Ensemble Size in a Tactile Communication System, S. B. Thesis, Department of Electrical Engineering, M.I.T., June 1963.

XXIV. STROBOSCOPIC RESEARCH

Prof. H. E. Edgerton
D. G. Kocher
J. A. McMorris II

A. MULTIPLE-SPARK LIGHT SOURCE AND CAMERA FOR SCHLIEREN AND SILHOUETTE PHOTOGRAPHY

This report summarizes a paper that is to be presented at the Eighth Annual Technical Symposium of the Society of Photographic Instrumentation Engineers to be held in Los Angeles, California, August 6, 1963.

A multiple-spark light source and camera have been devised by utilizing the Cranz-Schardin optical method for schlieren and silhouette photography. Each light source is a spark in air, with an effective duration of 10^{-7} second and a peak light of 24,000 candle power. The interval between successive sparks is independently adjustable from 10^{-5} to 10^{-7} sec, thereby permitting a variable picture rate to be used in a single sequence of photographs.

The equipment contains 10 spark light sources and camera lenses; thus a 10-picture sequence of high-speed phenomena is obtained. Schlieren and silhouette photographs demonstrating the performance of the equipment will be presented at the meeting.

A more complete report on this work will appear in our next quarterly report.

D. G. Kocher