UNCLASSIFIED

AD NUMBER

AD261909

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; 29 AUG 1961. Other requests shall be referred to the Air Force Research and Development Command, Washington, DC 20330.

AUTHORITY

USAEC ltr 15 Jun 1967

THIS PAGE IS UNCLASSIFIED

UNCLASSIFIED

AD 261 909

Reproduced
by the

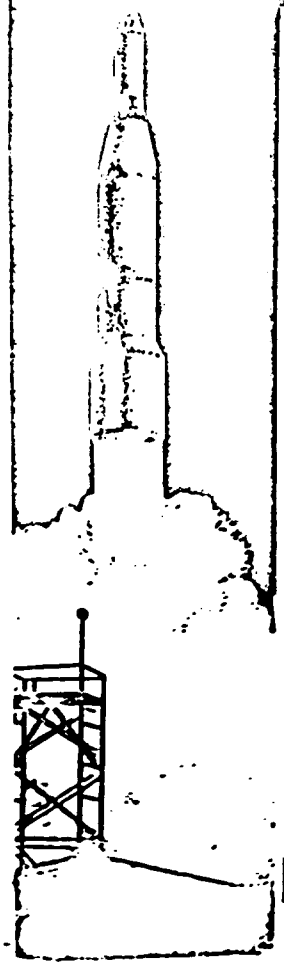
ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

AD 619 962
AS AD NO.
XEROX



4TH NATIONAL CONVENTION ON MILITARY ELECTRONICS

NO OTS.

1960 CONFERENCE PROCEEDINGS

JUNE 27, 28, 29, 1960
SHERATON PARK HOTEL
WASHINGTON, D.C.

ASTIA
AUG 29 1961

Best Available Copy

sponsored by
PROFESSIONAL GROUP on MILITARY ELECTRONICS · IRE

PUBLICATIONS COMMITTEE OF THE 1960 CONFERENCE PROCEEDINGS

Publisher

B. J. Goldfarb (*Chairman*)
Air Arm Division
Westinghouse Electric Corporation
Baltimore, Md.

Publication Distribution

George E. Martin
Weapon Control Department
Electronics Division
Westinghouse Electric Corporation
Baltimore, Md.

NOTE: Any comments or questions regarding specific papers requests for additional information, or individual reprints of these papers, should be forwarded directly to the respective authors.

Order additional copies and back issues from:

PROFESSIONAL GROUP ON MILITARY ELECTRONICS
INSTITUTE OF RADIO ENGINEERS
1 East 79th Street
New York 21, New York

1960 Conference Proceedings	$5.00
1959 Conference Proceedings	$4.00
1958 Conference Proceedings	IRE Members $3.00 Non-IRE Members $5.00
1957 Convention Record	IRE Members $3.00 Non-IRE Members $5.00

Printed by McGregor & Werner, Inc., Washington, D. C.

1960

CONFERENCE PROCEEDINGS

4th NATIONAL CONVENTION ON MILITARY ELECTRONICS



June 27, 28, 29, 1960
Washington, D. C.

– Sponsored by –

PROFESSIONAL GROUP ON MILITARY ELECTRONICS
INSTITUTE OF RADIO ENGINEERS

NATIONAL CONVENTION
PROFESSIONAL GROUP ON MILITARY ELECTRONICS

Military electronics is now at the very center of the U. S. defense effort. Its importance, and the urgency of its commitments, can hardly be overestimated if we consider all the factors that compose the historical situation we are living with, and shall live with in the years just ahead.

At the same time, the national significance of military electronics reinforces the drive toward constant development of new techniques, materials, processes, equipment, systems, and design ideas. As a result, military electronics is always on the frontier of the entire electronics industry. The contributions of military electronic engineers in their professional work are second to none in advancing the state of the art.

Yet this is a relatively new field. We are holding only the fourth of our national conventions as a professional group. The fact that we are a young member of the IRE family attests to the dynamism that has characterized the growth of our professional specialty.

So we may legitimately feel a deep satisfaction that this convention is important, because of its national significance; valuable, because of the technical nature of the work in its field; and progressive, because of the newness of our objects of study. The work of this convention will in itself be a contribution to scientific progress—as will the work of our local chapters across the country.

It is a pleasure to welcome all who are attending our fourth national convention. We look forward to a rewarding series of meetings and another year of growth and achievement afterward.

ROBERT L. CRAWFORD, Convention President

INTRODUCTION

The 4th National Convention on Military Electronics, sponsored by the Professional Group on Military Electronics of the Institute of Radio Engineers, has authorized publication of these PROCEEDINGS containing all the unclassified papers presented at the Convention. Also listed are the titles and authors of the classified papers which were presented under the sponsorship of the Air Research and Development Command, United States Air Force.

The response to our call for papers was gratifying. More than 400 abstracts were submitted for consideration. This year our convention has outgrown one hotel and we have had to use two. Our technical program is running 5 by 5. Five concurrent sessions were held in each of the five periods. With from 4 to 6 papers in each session a total of over 135 papers were presented. The Committee regrets that it was impossible to use all of the many very good abstracts submitted. The Committee offers its sincere thanks to those authors contributing to the Program. We recognize that the success and value of the convention rests mainly upon the efforts of the authors.

Those papers which were not available by the publisher's deadline date are being presented in the abstract form. Space is left on the page to allow members of the audience to take notes on the presentation. In some cases, the authors will have printed copies of their papers to distribute. The text and illustrations are reproduced as received from the authors. The papers are not edited for technical content so that the responsibility for the content of the papers published in this PROCEEDINGS rests upon the authors. Any comments or questions regarding the papers, requests for additional data, or individual reprints of these papers, should be forwarded directly to the respective authors. Requests for copies of individual classified papers should be made to the authors through appropriate channels in accordance with security regulations.

As Chairman of the Technical Program Committee I wish to take this opportunity to express my sincere thanks and appreciation to the individual Committee Members, Advisors and Moderators who assisted in planning and carrying out the Technical Program. The interest and support of the many cosponsoring companies, listed in the back pages of this publication, has helped toward making it possible to publish it and to provide each registrant with a copy without additional charge.

CRAIG M. CRENSHAW, Chairman
Technical Program Committee

PGMIL THROUGH THE YEARS

The successful organization and growth of the IRE Professional Group on Military Electronics over the past five years is an achievement that all members of the Group should know about.

The decision to organize the Group was made at a meeting during the 1955 IRE National Convention in New York City. Dr. Ernst Weber (past President of IRE), Mr. Lawrence G. Cumming and Mr. William M. Richardson met with Captain C. L. Engleman, USN (Ret.) who presented a written report of his study on the "pros and cons" of forming a "Professional Group on Military Electronics Operations." As a result of this meeting, Captain Engleman was asked to undertake the task of organizing the group.

The petition to form the new group on Military Electronics signed by interested IRE members was approved by the IRE Committee on Professional Groups the following September and by the IRE Executive Committee on 6 October 1955.

The first meeting of the charter National Administrative Committee for PGMIL was held on 12 December 1955, in Washington, D. C. The Committee consisted of the following eighteen members:

Maj. Gen. F. L. Ankenbrandt, USAF
Dr. James Q. Brantley, Jr.
Capt. Wm. I. Bull, USN
Mr. Carl R. Busch, (Secretary)
Mr. M. Barry Carlton (deceased)
Mr. Ralph L. Clark
Brig. Gen. Earle F. Cook, USA
Capt. C. L. Engleman, USN (Ret.) (Chairman)
Dr. Howard Engstrom (Treasurer)
Rear Adm. Fritz Furth, USN (Ret.)
Col. Gordon Gould, USAF (Vice Chairman)
Dr. Harvey Hall
Dr. George L. Haller
Mr. John E. Keto
Capt. David S. Little
Mr. William M. Richardson
Mr. Edwin A. Speakman
Brig. Gen. F. F. Uhrhane, USA

The approved Constitution spelled out the scope of the Group's activities as follows:

> "The field of interest of this group shall be concerned with the electronic sciences, systems, activities and services germane to the requirements of the Military . . .
>
> "This group will aid other Professional Groups of the IRE in their liaison with and services to the Military through joint meetings and activities, and by other appropriate means always keeping alert to the specific fields of interest of other Professional Groups."

The very next IRE National Convention (1956) found PGMIL very instrumental in providing the speaker for the notable "Highlight Session" reported on "Project Vanguard."

The first National IRE PGMIL Convention was held 17-19 June 1957 at the Sheraton Park Hotel in Washington, D. C. Of over 300 abstracts submitted, the program included some 80 technical papers. Classified sessions sponsored by branches of the Department of Defense were held in the Department of Commerce Auditorium. The banquet speaker was Lt. Gen. C. S. Irvine, USAF, and the luncheon speaker, Rear Adm. W. F. Raborn, USN. Some 100 military and industrial exhibits were on display.

In March 1957 the first issue of the Transactions of the PGMIL was published. The more frequent newsletter, MIL-E-GRAM, had made its appearance in October 1955.

By the end of June 1957, PGMIL boasted of some 3300 members with 12 chapters from Los Angeles to Long Island and Buffalo – Niagara to Central Florida. It had attained eighth place membership-wise of the 24 professional groups of IRE.

Rear Adm. W. E. Cleaves, USN (Ret.), who had served so ably as President of the 1957 PGMIL Convention, was elected Chairman of the PGMIL National Administrative Committee for 1957–1958. He advocated even stronger organizational and publishing policies in line with the favorable financial status of the Group.

The Second National Convention of PGMIL was held June 16-18, 1958 in Washington, D. C. Brig. Gen. W. B. Larew, USA (Ret.) served as President, Mr. W. M. Holaday as the banquet speaker, and Dr. S. W. Herwald was speaker for the luncheon.

For the year 1958-59, Mr. Edwin A. Speakman of RCA was elected National Chairman of PGMIL. Under his leadership membership expanded to over 4000. The PGMIL Transactions were revised to feature a "guest editor" for each edition, ensuring a continued high level of authorship for its contents. Mr. Speakman appointed Lt. Gen. William Kepner, USAF (Ret.) to be President of the Third National PGMIL Convention, which displayed a healthy growth over the two previous ones.

In March 1959 Mr. Speakman appointed Mr. B. J. Goldfarb, Westinghouse Electric Company, Air Arm Division to be Editor of MIL-E-GRAM, the Newsletter of the Professional Group on Military Electronics. He still continues the task of gathering interesting topical notes from the taciturn membership.

For the year 1959-60, Mr. Henry Randall of the Office of the Director of Defense Research and Engineering was elected National Chairman. The most significant innovation during this year has been the initiation of an additional national meeting.

In response to the demand from the West Coast PGMIL membership a Mid-Winter Convention was held in Los Angeles in February with Dr. L. C. Van Atta as President of the Convention. Over 2000 people participated in this attempt to serve the geographic areas of the country more uniformly. The operation was so well received that it will surely be repeated.

Dr. James Q. Brantley, Jr., and Dr. Donald R. Rhodes, both of Radiation, Inc., have served as Co-Editors of the PGMIL Transactions from the beginning. During Dr. Brantley's recent severe illness, Dr. Rhodes has carried the full responsibility.

Mr. R. H. Cranshaw was appointed Chairman of the 4th National Convention, which has begun to be called Mil-E-Con. Those of you who have attended these meetings are aware how this meeting exceeds the previous ones. The greater number of superior technical papers submitted and the numerous requests from exhibitors have created the pressure which has been met by the expanded operations in 1960.

4th NATIONAL CONVENTION
PROFESSIONAL GROUP ON MILITARY ELECTRONICS

CONVENTION ADVISORS

Dr. T. Keith Glennan	Administrator, National Aeronautics and Space Administration
Dr. Herbert F. York	Director of Defense R-E, Department of Defense
Admiral Arleigh A. Burke, USN	Chief of Naval Operations, Department of the Navy
Lieutenant General Arthur G. Trudeau, USA	Chief of Research and Development, Department of the Army
Lieutenant General Roscoe C. Wilson, USAF	Deputy Chief of Staff for Development, U. S. Air Force
Vice Admiral John T. Hayward, USN	Deputy Chief of Naval Operations for Development
Lieutenant General B. A. Schriever, USAF	Commander, Headquarters, Air Research and Development Command
Lieutenant General William E. Kepner, USAF (Ret.)	Chairman of the Board, Radiation, Inc.
Henry Randall	Office of Electronics, Office of the Director of Defense (R-E)
John E. Durkovic	Aeronautical Radio, Inc.

IH MILITARY ELECTRONICS
L INSTITUTE OF RADIO ENGINEERS, INC.

CONVENTION COMMITTEE

Convention President: Robert H. Cranshaw
Manager, Advanced Development Sales, Light Military Electronics Division, General Electric Company, 831 Broad St., Utica, New York

Convention Secretary: Joseph Avenasik, Jr.
U. S. Naval Research Laboratory, Washington, D. C.

EXHIBITS COMMITTEE

Chairman: L. D. Whitelock
Bureau of Ships, Washington, D. C.

R. C. Carpenter, Bureau of Ships, Washington, D. C.

E. P. Homburg, Sylvania Electric Products, Inc. Washington, D. C.

REGISTRATION COMMITTEE

Chairman: Jack M. Carter
Jansky and Bailey, Inc.
3629 M Street, N.W., Washington, D. C.

Thomas L. Sauter, Jansky and Bailey, Inc., 1339 Wisconsin Avenue, N.W., Washington, D.C.

Miss E. Callahan, Jansky and Bailey, Inc., 1339 Wisconsin Avenue, N. W., Washington, D. C.

ARRANGEMENTS COMMITTEE

Chairman. William H. Hulse
Asst. Eastern Regional Defense Manager
Westinghouse Electric Corporation
1625 K Street, N.W., Washington, D. C.

J. C. Lingle, Chief, Special Projects Office, Hq ARDC, RDA-3, Andrews Air Force Base, Washington, D. C.

Maj. J. K. Stein, USAF, Asst. Chief, Installation Security Branch Hq ARDC, RDISF, Andrews Air Force Base, Washington, D. C.

Paul Adams, District Sales Manager, Light Military, General Electric Company, Wyatt Building, 777 14th Street, N.W., Washington, D. C.

Mrs. Catherine Latimer, Installation Security Branch, Hq ARDC, RDISF, Andrews Air Force Base, Washington, D. C.

FINANCE COMMITTEE

Chairman: Jesse H. Adams, Engleman and Company, Inc., 2480 16th St., N.W., Washington, D.C.

William Graves, Hoye, Smith, Graves and Company, 1145 19th Street, N.W., Washington, D. C.

CONVENTION COMMITTEE - Continued

PUBLIC RELATIONS COMMITTEE

Chairman: Charles DeVore
CREI Atomics
3224 16th Street, N.W.
Washington, D. C.

Charles Garland, Office of Naval Communications, Washington, D. C.

Herbert H. Rosen, Deputy Public Information Officer, National Aeronautics and Space Admin., 1520 H Street, N.W., Washington, D. C.

TECHNICAL PROGRAM COMMITTEE

(Committeemen are also Technical Session Moderators)

Chairman: Dr. Craig M. Crenshaw
Department of the Army
Office of the Chief Signal Officer
Research and Development Division
Washington, D. C.

D. J. McLaughlin, Committee Advisor, U. S. Naval Research Laboratory, Washington, D. C.

Dr. Richard A. Weiss, Scientific Advisor, Office Chief Research and Development, U. S. Army, Washington, D. C.

Dr. Philip Newman, Chief, Propagation Laboratory, Air Force Cambridge Research Center, 230 Albany Street, Cambridge, Mass.

Dr. Joseph A. Boyd, Director, Project Michigan, University of Michigan, Ann Arbor, Mich.

Wilbur S. Hinman, Jr., Technical Director, Diamond Ordnance Fuse Laboratories, Ordnance Corps, U. S. Army, Washington, D. C.

Dr. Hans K. Ziegler, Chief Scientist, U. S. Army Signal Engineering Labs., Fort Monmouth, N. J.

Ralph I. Cole, Manager, Military Projects Planning, Melpar, Inc., 3000 Arlington Boulevard, Falls Church, Va.

G. Ross Kilgore, Manager, Applied Research Department, Westinghouse Electric Corp., P. O. Box 746, Baltimore, Md.

Dr. William J. Otting, Jr., Director, Physical Sciences, Hq Air Force Office of Scientific Research, Washington, D. C.

William S. Marks, Jr., Committee Advisor, Consultant, Zenith Radio Corporation

Robert F. Brady
Chief, Communications Branch
Research and Development Division
Office of the Chief Signal Officer
Department of the Army
Washington, D. C.

Responsibility for the contents of papers published in the Conference Proceedings of the Professional Group on Military Electronics, IRE, rests upon the authors. Statements made in papers are not binding on the IRE or its members.

TABLE OF CONTENTS

Session 1.1 RECONNAISSANCE AND RANGING (Confidential)—Sponsor: Air Research and Development Command. (Moderator: Dr. Joseph A. Boyd, Director, Willow Run Laboratories, Project Michigan, University of Michigan, Ann Arbor, Michigan)

Experimental Evaluation of a Diversity Radar
Philip W. Crist, Airborne Instruments Laboratory, Melville, N. Y.

Radar Density Related to Tactical Troop Concentrations
R. T. Stefancik and G. A. Pitsenbarger, Sylvania Electronic Defense Laboratory, Mountain View, Calif.

A Submarine Simulator Device Program
Robert H. Dickman, U. S. Naval Training Device Center, Port Washington, N. Y.

Identification and Evaluation of Magnetic Field Sources Associated with Magnetic-Anomaly Detector Equipped Aircraft
Paul Leliak, The Martin Company, Baltimore 3, Maryland

Session 1.2 SATELLITE ELECTRONICS (Moderator: Dr. Hans K. Ziegler, Chief Scientist, U. S. Army Signal Research and Development Laboratory, Fort Monmouth, N. J.)

A Satellite Microwave Telemetry Oscillator Using Traveling-Wave Tube Techniques 451
L. A. Roberts, H. R. Johnson, Watkins-Johnson Company, Palo Alto, Calif.

Application of Microminiaturization Concepts to Space Guidance Computers 1
Edward Keonjian, American Bosch Arma Corporation, Hempstead, N. Y.

Criteria for the Optimum Design of Active Satellite Communication Systems 6
Albert R. Giddis, Philco Western Development Laboratory, Palo Alto, Calif.

Satellite Ionosounder 13
Samuel Horowitz, AF Cambridge Research Center, Bedford, Mass. and Leonard Humphrey, General Electric Company, Ithaca, N. Y.

Satellite Relicbility Achieved Through Comprehensive Environmental and Functional Testing 455
John A. Chambers, U. S. Army Ordnance Missile Command, Redstone Arsenal, Ala.

A Compact UHF Diplexer for Applications Involving Rockets or Satellites 461
Sam E. Parker, Hughes Aircraft Company, Culver City, Calif.

Session 1.3 MICROWAVE DEVICES AND TECHNIQUES (Moderator: G. Ross Kilgore, Westinghouse Electric Corporation, Baltimore, Md.)

Status of Ultra-Low-Noise Traveling-Wave Tubes and Beam-Type Parametric Amplifiers 466
K. Kotzebue, B. P. Ismelsen and G. E. St. John, Watkins-Johnson Company, Palo Alto, Calif.

Bulk Semiconductor Devices for Microwave Applications 472
H. Jacobs, F. A. Brand, M. Benanti, J. Meindl, and R. Benjamin, U. S. Army Signal Research and Development Lab., Fort Monmouth, N. J.

Phase Stable Limiting I-F Amplifiers Using Beam Deflection Tubes 71
E. R. Wingrove, Jr., General Electric Company, Syracuse, N. Y.

Applications of Traveling-Wave Tubes to Microwave Circuits 24
G. E. Austin, Sylvania Electronic Defense Laboratory, Mountain View, Calif.

Recent Electronic Scanning Developments 30
J. P. Shelton and K. S. Kelleher, Aero Geo Astro Corporation, Alexandria, Va.

Practical Stripline Component Design 589
Virginia T. Norwood, Hughes Aircraft Company, Culver City, Calif.

Session 1.4 INSTRUMENTATION – GENERAL (Moderator: Dr. William J. Otting, Jr., Director of Physical Sciences, Air Force Office of Scientific Research, Washington 25, D. C.)

The Pulsed Light Theodolite 35
Lee A. Jay, U. S. Army Electronic Proving Ground, Fort Huachuca, Ariz.

Multiple High PRF Ranging 37
David R. Mooney and Wm. A. Skillman, Westinghouse Electric Corp., Baltimore 3, Md.

Recent Achievements in Missile-Borne Magnetic Recorders 41
Mark M. Siera, Lockheed Aircraft Corporation, Palo Alto, Calif.

Hit Indicator Techniques for Direct Fire Weapons 49
Herbert Chaskin, U. S. Naval Training Device Center, Port Washington, N. Y.

Effects of Atmospheric Pollutants on Electronic Equipment 49
Herbert C. McKee, Southwest Research Institute, San Antonio, Texas

A Delay-Line Synthesized Filter Bank with Electronically Adjustable Impulsive Response 498
H. J. Bickel, and E. Brookner, Federal Scientific Corp., New York, N. Y.

Session 1.5 NOISE EFFECTS ON PRECISION AND DATA (Moderator: M. D. McFarlane, Robertshaw-Fulton Controls Company, Anaheim, Calif.)

0.1% Accuracy Variable Speed Control System 556
Louis Goodman, Fairchild Camera & Instrument Corporation, Syosset, L. I., N. Y.

Pulse Operation of D-C Servo Motors for Lower Thresholds .....
D. J. Salesimer and William E. Yoakum, Guided Missile Agency, Redstone Arsenal, Ala

A Mathematical Analysis of Transients Caused by AGC Reset in Line 54
Switching Amplifier as Used in AN/MSQ-18 (Missile Monitor) Equipment
SP-4 John M. Dugan and 1st Lt. R. A. Perry, U. S. Army Air Defense Board, Fort Bliss, Texas

An Analogue Computer for Separating Evoked Physiological Potentials from Background Noise 57
Walter Kropil, Robert Robinson and John C. Armington, D. I. Tepas, Walter Reed Army Institute of Research, Washington 12, D. C.

Precision Frequency Measurement of Noisy Doppler Signals 61
William A. Dean, Ballistic Measurements Laboratory, Aberdeen Proving Ground, Md.

Output Signal-to-Noise Characteristics of Correlators 70
Bruce P. Mayo, General Electric Company, Syracuse, N. Y. and
David K. Cheng, Syracuse University, Syracuse, N. Y.

Session 2.1 COMMUNICATIONS AND DATA HANDLING (Confidential)—Sponsor: Air Research and Development Command (Moderator: Mr. Robert F. Brady, Research and Development Division, Office of the Chief Signal Officer, Department of the Army, Washington 25, D. C.)

Engineering Analysis of Qualitative Data to Provide a Missile System for Simulation and Vulnerability Studies
N. Johnson, Sylvania Electronic Defense Laboratory, Mountain View, Calif

SPASUR Automatic Digital Data Assembly System
Part I—The Digital Data Transmission Problem
Part II—Description of the Digital System
Mr. W. B. Poland, Jr., U. S. Naval Research Laboratory, Wash., D. C. and M. S. Maxwell, J. Pinker, U. S. Naval Weapons Laboratory, Dahlgren, Va.

Simulation by Interpretation
Miss Ruth Baldwin, F. W. Fina, Jr., and J. J. Wolf, Burroughs Corporation, Paoli, Pa.

A Transistorized Digital Range Unit
R. M. Lucas, Bell Telephone Laboratories, Inc., Whippany, N. J.

Evaluation of Video & IF MTI Cancellation Techniques
E. C. Nordell, General Electric Company, Dewitt, N. Y.

Session 2.2 COMMUNICATIONS I (Moderator: Captain William L. Bull, USN, Bureau of Ships, Main Navy Building, Washington, D. C.)

Military Applications for Speech Compression Techniques 73
A. J. Strassman and K. C. Stockhoff, Hughes Aircraft Company, Los Angeles 45, Calif.

Instrumentation Used for Ionosphere Electron Density Measurements 497
William J. Cruickshank, Ballistic Research Laboratory, Aberdeen Proving Ground, Md.

Electron Density Measurements in Hypersonic Projectile Trails .....
R. Scott Hebbert, U.S. Naval Ordnance Laboratory, White Oaks, Silver Spring, Md.

Topology Engineering of Communication Networks 82
Dr. Kurt Ikrath and Charles C. Comstock, U. S. Army Signal Research & Development Laboratory, Ft. Monmouth, N. J.

Frequency Selection 600
Howard R. Smith, Sierra Vista, Ariz.

The Economic Design of Radio Communication Systems by Matching the Message Urgency to the Fading Conditions .....
Leang P. Yeh, Page Communications Engineering, Inc., Wash., D. C.

Session 2.3 RELIABILITY I (Moderator: Charles A. Strom, Jr., Directorate of Communications, RADC, Griffiss AFB, N. Y.)

Serviceability: Complement to Reliability 98
Richard H. Wilcox and CDR Vance R. Wanner, Office of Naval Research, Washington 25, D. C.

Reliability Anatomy for System Design Engineers 103
E. S. Winlund, Sunnberg-Carlson Company, San Diego, Calif.

Performance of a One-Unit System 111
Richard E. Barlow and Larry C. Hunter, Sylvania Electronic Defense Laboratory, Mountain View, Calif.

Statistical Pitfalls for the Reliability Engineer 117
G. H. Beckhart, Radio Corporation of America, Moorestown, N. J.

The Use of IBM Cards to Predict, Control, and Measure Reliability of a Missile Electronics Unit 120
George F. Dolan, Hughes Aircraft Company, Culver City, Calif.

System Improvement Through Failure Effect and A Reliability Model 128
George E. Unruh, Paramount Electronics, Inc., Hicksville, N. Y.

Session 2.4 ANTENNAS (Moderator: Lt. Colonel George F. Watkins, Air Research and Development Command, Washington 25, D. C.)

Ferromagnetic Antennas 135
Ronald New, American Electronic Laboratories, Inc., Lansdale, Pa.

Effect of Antenna Phase Pattern on Doppler System Operation 503
H. S. Rothman and W. E. Scharfman, Stanford Research Institute, Menlo Park, Calif.

A Generalized Analysis of Electronic Antenna Beam Steering 140
P. D. Kennedy, Lockheed Aircraft Corporation, Sunnyvale, Calif.

The Shroud Antenna for High Speed Missiles 510
Victor W. Richard, Ballistic Research Laboratory, Aberdeen Proving Ground, Md.

Ring Arrays 153
W. D. Nelson, General Electric Company, East Dewitt, N. Y.

A Flush-Mounted VHF Telemetry Antenna with Hemispherical Coverage 242
Ronald C. Payne and Parker Painter, Jr., Dynatronics, Inc., Orlando, Fla.

Session 2.5 RANGING, TRACKING AND RECONNAISSANCE (Moderator: Herbert Goldway, U. S. Army Combat Surveillance Agency, Arlington 1, Va.)

Continuous Wave Range System 517
S. C. Lanza, Philco Corporation, Palo Alto, Calif.

An Experimental Study of Monopulse Technique for Ground Clutter Discrimination . 521
Shuzuan Y. Chang and Vincent Scabilito, U. S. Army Ordnance, Frankford Arsenal, Philadelphia 27, Pa.

Tracking Research at the U. S. Naval Training Device Center 157
Gene Michels, U. S. Naval Training Device Center, Port Washington, N. Y.

On the Tracking and Geodetic Potentialities of a Doppler Rate Measuring System 161
Duance C. Brown, RCA Service Company, Patrick Air Force Base, Fla.

Instrumentation Error Analysis of the AMR Missile Tracking Systems 335
Bruce U. Glass, RCA Service Company, Patrick Air Force Base, Fla.

Calculation of Measurement Errors by a Statistical Technique 168
Larry G. Larson, Philco Corporation, Palo Alto, Calif.

Session 3.1 GUIDANCE AND SPACE TECHNOLOGY (Confidential)–Sponsor: Air Research and Development Command (Moderator: Colonel B. S. Pulling, WADD, Wright-Patterson Air Force Base, Ohio)

Integrated Design of Antennas for Ballistic Missiles and Space Vehicles
D. A. Alsberg, Bell Telephone Laboratory, Inc., Whippany, N. J. and H. W. Redlien, Wheeler Laboratory, L. I., N. Y.

The Space Surveillance System
Dr. C. E. Cleeton and Roger L. Easton, U. S. Naval Research Laboratory, Washington 25, D. C.

Orbit Determination for Passive Satellite Detection
R. B. Patton, Jr., Ballistic Research Laboratory, Aberdeen Proving Ground, Md.

Evaluation of Reliability Prediction Techniques of the Electronics System of the Falcon Missile
Frank A. Barta, Hughes Aircraft Company, Culver City, Calif.

Session 3.2 DATA HANDLING (Moderator: Mr. D. J. McLaughlin, U. S. Naval Research Laboratory, Washington, D. C.)

Factors Influencing the Research and Development of New Computer Programming Techniques Required for Mechanization of Machine Learning 173
Dr. Robert E. Smith, Control Data Corporation, Minneapolis 15, Minn.

Pattern Recognition 179
Jos. W. Brouillette and Charles W. Johnson, General Electric Company, Syracuse, N. Y.

New Techniques in Residual Arithmetic 183
Michael R. Levine and Julius Metz, American Bosch Arma Corporation Hempstead, N. Y.

A Note on the Applicability of Error Correcting Codes 190
J. E. Palmer, Radio Corporation of America, Camden 2, N. J.

A Real Time Telemetry Data Transmission System 194
H. E. Rennacker, Collins Radio Company, Burbank, Calif.

Computer Controlled Automatic Diagnostic and Checkout System for Field Use ... 200
Raymond J. Brachman, Frankford Arsenal, Philadelphia, Pa.

Session 3.3 SPECIAL ELECTRICAL COMPONENTS (Moderator: Colonel R. E. Kimball, Office of Chief Research and Development, Army Research Office, Arlington, Va.)

Some Aspects of Tunnel Diode Applications 203
Thomas O. Krueger, U. S. Army Signal Research and Development Laboratory, Fort Monmouth, N. J.

The "Rayistor" an Electrical Transformer Using Optical Coupling 212
James C. Davis, Jr., Raytheon Company, Bedford, Mass.

Thin Film Components Based on Tantalum 214
R. W. Berry and N. Schwartz, Bell Telephone Laboratory, Inc., Murray Hill, N. J.

Thermoelectric Generators for Short Term Application 218
Samuel R. Hawkins, Lockheed Aircraft Corporation, Sunnyvale, Calif.

New Developments in the Field of Military Quartz Crystals 225
G. K. Guttwein, U. S. Army Signal Research and Development Laboratory, Fort Monmouth, N. J.

Session 3.4 RADAR (Moderator: Harry Mayer, General Electric Company, Ithaca, N. Y.)

Certification of Command Guidance Missile Systems 220
C. M. Redman, Headquarters, White Sands Missile Range, N. M.

System Evaluation of Low Noise Radar Sensitivity 234
S. Charton and G. Ver Wys, Radio Corporation of America, Moorestown, N. J.

The Implementation of the Integrated Mapping System 563
Arthur Fein, Fairchild Camera and Instrument Corporation, Syosset, L. I., N. Y.

The Effect of Field Wire Stability on the Maximum Length of Loop 147
N. W. Feldman and G. P. Tripp, Headquarters, U. S. Army Signal Research and Development Laboratory, Fort Monmouth, N. J.

Limitations of Angular Radar Resolution 249
Dr. Ewald Eichler, U. S. Army Ordnance, Frankford Arsenal, Philadelphia 37, Pa.

Internal Ballistic Measuring System .....
Leo Adelson, Picatinny Arsenal, Dover, N. J.

Session 3.5 SIMULATION—GENERAL (Moderator: Captain Thomas M. Adams, USN, Bureau of Naval Weapons, Main Navy Building, Washington, D. C.)

Space Simulation with High Gas Release Rates 253
W. W. Balwanz and J. M. Singer, Naval Research Laboratory, Washington 25, D. C.

Tank vs. Tank Synthetic Gunnery Trainer 258
Irwin Friedland, U. S. Naval Training Device Center, Port Washington, N. Y.

Mathematical Models of Multiple-Gimbal Systems 265
Dr. Azriel Rosenfeld, Budd Lewyt Electronics, Inc., Long Island, N. Y.

Optimization of Test Systems 272
John C. O'Brien, Cooper Development Corporation, Monrovia, Calif.

Optimum Search Routines for Automatic Fault Location .....
Sidney L. Firstman, The Rand Corporation, Santa Monica, Calif. and Brian Gluss, Armour Research Foundation of Illinois, Chicago 16, Ill.

Session 4.1 ELECTRONIC GENERATION, SWITCHING AND RADIATION (Confidential)—Sponsor: Air Research and Development Command (Moderator: Frank T. Mitchell, Jansky & Bailey, Alexandria, Va.)

A Very High Gain Experimental Millimeter Foster Scanner
C. A. Hacking, I-T-E Circuit Breaker Company, Philadelphia 34, Pa.

A Special Purpose Microwave Switch for Anti-jam Operation of Conical Scanning Radars
S. D. Schreyer and G. Klein, Westinghouse Electric Corporation, Baltimore 3, Md.

A Periodically-Focused 10 KW Microwave Traveling-Wave Amplifier
O. T. Purl and K. W. Slocum, Watkins-Johnson Company, Palo Alto, Calif.

Electromagnetic Radiation in Sea Water
E. J. Hilliard, U. S. Naval Underwater Ordnance Station, Newport, R. I.

A Radar Technique Using an Electro-Optical Two-Dimensional Filter
Part I — Principles of Operation
Part II — An Experimental Model Employing a Delay-Line Light Modulator
Louis Lambert, Moses Arm, and Isaac Weissman, Columbia University in the City of New York, New York 27, N. Y.

Session 4.2 COMMUNICATIONS II (Moderator: William S. Marks, Jr., Winter Park, Fla.)

Impact of the Recent International Radio Conference, Geneva, on USAF Communications-Electronics Programs 278
Carl W. Loeber, Department of Air Force, Washington, D. C.

Communication by Re-radiation from Chaff 542
B. V. Blom, Benson, Ariz.

The Use of Faraday Rotations in Prediction of Ionospheric Disturbances 282
Hallock S. Marsh and R. J. Cormier, Air Force Bambridge Research Center, Bedford, Mass.

Digital Battlefield Communications 289
William C. Slagle, Stromberg-Carlson, San Diego 12, Calif.

Session 4.3 RELIABILITY II (Moderator: Frank H. Clarke, Texas Instruments, Inc., Washington, D. C.)

Participation in Various Data Exchange Programs 294
Stanley L. Pollock, U. S. Naval Ordnance Laboratory, Corona, Calif.

Reliability of a Parallel System Considering Load Redistribution 302
C. H. Tseo and H. L. Leve, Hughes Aircraft Corporation, Culver City, Calif.

Understanding and Improving System Reliability and Maintainability Using Information in Engineering/Environmental Malfunction Data Samples 308
George H. Allen, Raytheon Company, Maynard, Mass.

The Reliability of Hermetically Sealed Equipment 315
W. B. Rossnagel, Kearfott Division of General Precision, Inc., Little Falls, N. J.

A Measure of Reliability and Information Quality in Redundant Systems 320
S. A. Rosenthal, H. Jaffe, and M. D. Katz, Sperry Gyroscope Co., Great Neck, N. Y.

Session 4.4 SYSTEMS ANCILLARY TO MISSILES (Moderator: Sol Levine, The Martin Company, Baltimore 3, Md.)

Instrumentation System for Three-Dimensional Tracking of Underwater Missiles ... 618
C. S. Soliozy and J. M. Formwalt, U. S. Naval Underwater Ordnance Station, Newport, R. I.

MATTS (Multiple Airborne Target Trajectory System) 584
Walter J. Zable, Cubic Corporation, San Diego 11, Calif.

Equipment Design Trends in Missile Scoring Devices 605
W. Ficklin, A. H. Maciszewski, J. J. Pakan and K. Ringer, A. R. F. Products, Inc., River Forest, Ill.

Airborne Instrumentation Systems Utilized in First and Second Generation Ballistic Re-entry Vehicles .....
Leroy E. Foster, General Electric Company, Philadelphia 4, Pa.

Flight Measurements on the JUPITER R&D Missile 547
C. T. N. Paludan, Huntsville, Ala.

Session 4.5 SIMULATION—ELECTRONIC (Moderator: Dr. Richard A. Weiss, Scientific Advisor, Office Chief Research and Development, U. S. Army, Washington 25, D. C.)

An Evaluation of Techniques in Land Mass Radar Simulation 568
Charles Colbert and George M. Dembo, Westgate Laboratory, Inc., Yellow Springs, Ohio

Electromagnetic Environment Simulation for System Trainers 331
Francis P. Cullen, William Helf, and John K. Scully, The Marquardt Corporation, Pomona, Calif.

The Development of a Dynamic Target and Countermeasures Simulator 338
Richard L. Norton, U. S. Army Signal Missile Support Agency, White Sands Missile Range, N. Mex.

Celestial Navigation 343
Gerard Jaquiss, U. S. Naval Training Device Center, Port Washington, N. Y.

A Synthetic Trainer for Tank Turret Crewmen 349
Thomas Mongello, U. S. Naval Training Device Center, Port Washington, N. Y.

Session 5.1 INSTRUMENTATION (Confidential)—Sponsor: Air Research and Development Command (Moderator: Charles A. Stec, Bureau of Ships, Navy Department, Washington 25, D. C.)

Electric Firing of Fully Combustible Ammunition
F. J. Dashmaw, Watervliet Arsenal, Watervliet, N. Y.

Transmission of Electromagnetic Waves Through An Ionized Medium in the Presence of a Strong Magnetic Field
T. P. Hurley, Boeing Airplane Company, Seattle 24, Wash.

The Exploitation of Millimeter Waves for Military Applications
Harold N. Tate, Hq, U. S. Army Signal Research & Development Laboratory, Ft. Monmouth, N. J.

A Millimeter Wave Radar System
George Howitt, Allen B. DuMont Laboratories, Inc., Clifton, N. J.

Increased Jamming and Reconnaissance Effectivity Through Polarization Diversity
E. F. Henry, Melpar, Inc., Falls Church, Va.

Session 5.2 SPACE TECHNOLOGY (Moderator: Dr. Dieter H. Schwebs, The Advanced Research Projects Agency DOD, Washington 9, D. C.)

Maintenance, Repair and Assembly in Space by Remote Means 355
John W. Clark, Hughes Aircraft Company, Culver City, Calif.

A Functional Description of Pioneer V 361
Paul F. Glaser, Space Technology Laboratories, Inc., Los Angeles, Calif.

Communication in Space by Deflected Sunlight 367
K. W. Otten, Wright Air Development Division, Wright-Patterson Air Force Base, Ohio

Optimum Capacitor Charging Efficiency for Space Systems 575
Drs. Philip M. Mostov, Jos. L. Neuringer and Mr. Donald S. Rigney, Republic Aviation Corporation, Farmingdale, L. I., N. Y.

Application of Inertial Techniques to Interplanetary Navigation 373
Milton J. Minneman, Republic Aviation Corporation, Farmingdale, L. I., N. Y.

Session 5.3 CAMERA DISPLAY DEVICES (Moderator: Ralph I. Cole, Melpar, Inc., Falls Church, Va.)

High Performance Camera Tubes 380
Sidney Gray, RCA Laboratories, Princeton, N. J.

High Speed Direct Electronic Printing Cathode Ray Tube 623
N. Fyler, D. Cone, R. Dorr, J. Wurtz, Litton Industries, San Carlos, Calif.

Extending the Dynamic Range of Camera Tubes Employing Return Beam Modulation 385
A. D. Cope and H. Borkan, Radio Corporation of America, Princeton, N. J.

Image Orbicon Tubes as Image Intensifiers 390
Nils Swanson and James Parton, U. S. Army Engineer Research and Development Laboratory, Fort Belvoir, Va.

Session 5.4 VULNERABILITY, GUIDANCE AND CONTROL (Moderator: Wilbur S. Hinman, Jr., Technical Director, Diamond Ordnance Fuze Laboratories, Ordnance Corps, U. S. Army, Washington 25, D. C.)

Loran-C Navigation System .....
William Dickinson, Jansky & Bailey, Inc., Washington, D. C.

A New Gyro for Autopilot Use 593
Sahag Dardarian, General Precision, Inc., Little Falls, N. J.

Lightweight Inertial Systems 396
Robert E. Marcille, Litton Industries, Beverly Hills, Calif.

Inertial Accelerometers—Their Nature, Character and Limitations 403
Martin Mauerer, General Precision Inc., Little Falls, N. J.

Jamming Effectiveness Instrumentation 413
Capt. G. H. Redwine and Charles H. Meyer, III, Rome Air Development Center, Griffiss AFB, N. Y.

Session 5.5 DATA HANDLING (Moderator: Henry L. Metz, CREI, Washington 10, D. C.)

High Speed Auto-Data System for Blast Studies 420
R. D. Jones and J. D. Smith, Sandia Corporation, Sandia Base, Albuquerque, N. Mex.

Data Handling for a Research and Development Test Stand 430
Thomas Wong and Robert L. Thomason, U. S. Naval Ordnance Test Station, China Lake, Calif.

The Handling of UHF Doppler Data 435
Douglas H. Parks, Radio Corporation of America, Patrick Air Force Base, Fla.

An Automatic Data Reduction Facility Combining Maximum Versatility and Speed 440
William R. Schumacher, U. S. Navy Underwater Sound Laboratory, New London, Conn. and H. Malcolm Wilkinson, Epsco, Inc., Cambridge, Mass.

The Digitron—A High Speed Data Display System 445
P. J. Meredith, D. J. Griffin, and F. A. Paulus, The Marquardt Corporation, Pomona, Calif.

APPLICATION OF MICROMINIATURIZATION CONCEPTS TO SPACE GUIDANCE COMPUTERS

By: Edward Keonjian, American Bosch Arma Corporation

Introduction

Modern digital computers for space guidance are very complex machines with many functional blocks which normally house thousands of individual electronic components. In addition, the computer must be capable of functioning in severe mechanical, climatical and nuclear environments.

If present day electronic components and packaging techniques were used, the size and weight of such a computer would become enormous.

By application of various novel concepts of microminiaturization to such a computer, a significant reduction in its size and weight could be achieved, combined with greater potential reliability.

These new and rapidly growing concepts have a variety of forms, depending on the technique used, and in general, are applicable to low power level electronic circuitry.

The presently existing concepts of microminiaturization can be divided into the following four basic approaches.

1. Integration of one or more individual components of the same type on flat substrates which in turn are integrated into 3-dimensional functional circuit modules (referred to as "micro-module concept").

2. Integration of various group of various components on a single substrate to form an entire functional circuit (referred to as "two-dimensional concept").

3. Formation of all the elements of a functional circuit on a single semiconductor substrate (referred to as "solid-state concept").

4. Formation of circuit function at microscopic levels, combining molecular particles to obtain any predetermined transfer characteristic (referred to as "microcircuitry" or "molecular circuitry").

It is also possible to combine various approaches to produce a specific circuit, if there are some advantages to be gained. However, the above mentioned approaches have already been established as basic for the present state of the art.

Component Density

Component density is an important factor which is directly associated with any concept of microminiaturization (though in type 3 and 4 approaches, "functional circuits density" would be a more proper term). Generally, it indicates how many components (or circuit functions) can be "squeezed" into a given volume without imperiling the normal operation of the circuit.

In integrated circuits, the components' density will depend largely upon the volumetric efficiency of individual components. The latter is defined as the ratio of the volume of the active element to the total volume of a component. In general, for most of the conventional components, volumetric efficiency is very small. For example, in a JEDEC 30 transistor case, which has a volume of approximately 0.08 cubic inches, the active element (semiconductor wafer) is in the order of 6×10^{-5} cubic inches. This means that the "workable" part of a transistor constitutes only $\frac{1}{1000}$th of its total volume! The balance is utilized for mounting, interconnecting and protecting the active element.

Similarly, the thin carbon or metal film in individual resistors is a very small percentage of the total volume used up even by miniature units. The same is true for diodes and for many types of capacitors.

Thus, the first step towards microminiaturization would be elimination of a large portion of this superfluous component "housing" by putting "stripped" components of an entire functional circuit on a very thin wafer substrate. In realization of this idea, the type 1 and 2 concepts of microminiaturization were born, offering a component density up to 10^6 components per cubic foot. However, taking into account the space occupied by interconnections and other assembling media, the final part density for an operational electronic equipment may vary between 2×10^5 and 6×10^5 components per cubic foot, depending on the type of circuit. This is approximately a 10:1 reduction in size and weight, compared to conventional equipments of the same function.

In microminiaturization concepts type 3 and 4, when components lose their individual identities, the merit of their density may be expressed by a term called "equivalent component density". It is based on the dimensions (volume) of a type 3 or 4 functional circuit, and the number of individual components that would be required to perform the same circuit function if it were made of those components. For example, a solid-state flip-flop measures $\frac{1}{4} \times 1/8 \times 1/32\ 10^{-3}$ cubic inches. If made of conventional components, the flip-flop would require 20 components. This number of components in the volume of 10^{-3} cubic inches, represents an "equivalent part density" in order of 30×10^6 parts per cubic foot.

Theoretically, the type 4 concept of microminiaturization offers an almost unlimited "equivalent parts density" for low level electronic circuits. However, the development of an adequate interconnection technique is of paramount importance for the full utilization of type 4 concept, as in all other concepts of microminiaturization.

Problem of Interconnections

The interconnections are not only wasteful from the point of view of the equipment volumetric efficiency, but what is even more important, they adversely affect the reliability of the equipment,

since every single interconnection should be considered as a potential source of unreliability, a potential source of a failure.

To illustrate the seriousness of this problem, one has to consider that a typical airborne computer may have as many as 40,000 interconnections in a logic circuit alone!

"We are already approaching a limit for complexity with an acceptable level of reliability", said Dr. D. E. Nobel of Motorola. "Continuing this line of development will lead us to the ultimate catastrophe - a system of maximum complexity which will never work".

Consequently, only by reducing the number of interconnections which will lead to equipment of less complexity, can higher reliability be attained.

For this reason the reduction of physical size of electronic components alone, without proper emphasis on their interconnections, cannot be of a great practical value to one concerned with the problems of building reliable microminiature electronic equipment for space application.

Development of proper interconnection technique should also make possible an easy "throw-away" type of maintenance; i.e., the interconnection should permit easy removal of an improperly functioning building block and its replacement with a new one. Thus the maintainability of the equipment is another important factor of microminiature electronic equipment.

Reliability and Redundancy

Extremely high reliability, which is required from individual electronic components and functional blocks of a space equipment, makes the cost so high that sometimes it is difficult to justify commercially. A remedy to this problem can be found in redundancy which may become practical in view of small size (and hopefully low cost) of microminiature devices. According to Dr. J. VonNeumann, the principle of redundancy is "used" even in the human brain, which is composed of individual brain cells of quite a low reliability. By using thousands of cells where perhaps one could have done the job, and a very sophisticated self-analysis mechanism that allowed the choice of the best operating cells, it was possible to achieve reliability factors of almost infinite magnitude.

Application of 2-D Concept

The "two-dimensional" concept of microminiaturization has been adapted as the most feasible approach for a space guidance computer of 1961. This concept is based on putting the "stripped" components of an entire functional circuit on a very thin wafer substrate. Since it is an almost two-dimensional package, the name "2-D" has been adopted for this approach. To obtain maximum part density, a number of these wafers are stacked for form a functional block.

Figure 1 shows a close-up of a 2-D amplifier, with deposited resistors and integrated capacitors and semiconductors. The dimensions of the circuit are 0.5" X 0.5" X 0.03".

In a more advanced form of 2-D concept, it is feasible to deposit all the components (except semiconductors), and their principle interconnections on a substrate, using a single metal, thus attaining a higher part density and higher reliability.

Following electrical tests, a number of 2-D circuits can be integrated into a larger functional building block. In the foreground of Figure 2 there are two such blocks, each of which is a computer full adder. Each block consists of seventeen individual circuits: seven amplifiers, eight AND gates and one flip-flop. The size of individual circuits is 0.5" X 0.03". The total number of components employed by the adder is eighty five: nine transistors, thirty-five diodes, thirty-seven resistors and four capacitors. The volume of the entire circuit, including all interconnections is 0.5 cubic inches. This corresponds to a circuit parts density of approximately 600,000 parts per cubic foot.

Protection of the individual circuit wafers as well as the subassemblies and circuit groups against severe environments is another important problem of microminiaturization. So far, most of the attempts to coat circuit components with a suitable protective substance did not give entirely satisfactory results. Therefore, as a compromise solution to this problem, the circuit subassemblies and functional blocks can be encapsulated in a thin walled, hermetically sealed plug. The plug can then be filled with an inert, high heat conductive gas. (See Figure 3).

The storage unit of the computer contributes a considerable amount to its volume and weight. Therefore, various methods have been studied which would lead toward microminiaturization of computer storage elements and assemblies. The thin film techniques permit storing a large amount of information in the minimum space. However, this technique is still in the experimental stage. Present plans involve the use of subminiature transfluxors, multiaperture ferrite cores with approximately 0.1" outside diameter. Figure 4 shows the wiring and packaging approach of such transfluxor storage assembly for approximately 30,000 bits of memory.

Figure 5 shows an exploded view of the complete assembly of the microminiature computer. The estimated weight is 12 pounds. The general view of the computer is shown in Figure 6.

Figure 7 illustrates the relative size reduction of typical computers that has been accomplished during the past five years and predicts the improvements of the next five years.

Conclusion

The major problem confronting the entire concept

of microminiaturization today is reliability.

To achieve high reliability, the system must be simple with a minimum number of interconnections or it has to use redundancy in design.

The 2-D Concept adapted for ARMA's Space Computer, is an approach which can be materialized in 1961-62.

The success of any advanced approach will depend on the development of processes for producing the circuit. Without proper emphasis on producibility, a discussion of the relative merits of various concepts of microminiaturization would be inconclusive and even misleading.

Acknowledgment

The author wishes to thank his associates in this program: E. Harmon, R. Laird, W. Lukac, J. Maguire, M. Schiller, I. A. Weintraub and V. Wohler.

References

1. J. T. Wallmark, "Design Considerations for Integrated Electronic Devices" Proc. I.R.E.; Vol. 48; March 1960.

2. W. W. Gartner, "Integrated Circuitry, Micro-Miniaturization and Modular Electronics"; Semiconductor Products, December 1959.

3. S. F. Danko, W. L. Doxey and J. P. McNaul, "The Micro-Module: A Logical Approach to Microminiaturization", Proc. I.R.E.; Vol. 47, May 1959.

4. T. A. Prugh, J. R. Mall and J. J. Doctor, "The DOFL Microelectronics Program", Proc. I.R.E.; Vol. 47; May 1959.

5. L. D. Shergalis, "Microelectronics - A New Concept in Packaging", Electronic Design, April 1959.

6. "Microminiature Designs Heading for Production"; Electronic Design, Dec. 23, 1959.

7. "Molecular Electronics", Electronic Industries, March 1960.

8. C. G. Childs, A. P. Haase, M. W. Hamilton and R. M. Hughes, "The Thermionic Integrated Micro-Module Program", The 1960 I.R.E. International Convention, New York, N.Y., March 1960.

9. J. S. Kilby and E. Keonjian, "Design of Semiconductor Solid-Circuit Adder", The 1959 Electron Devices Meeting, Washington, D.C.; October 1959.

10. J. S. Kilby; "Semiconductor Solid Circuits", Electronics, Vol. 32; August 7, 1959.

11. G. W. A. Dummer; "British Approaches to Microminiaturisation"; Electronics, January 1960.

12. E. Keonjian; "Microminiature Electronic Circuitry for Space Guidance", 1959 I.R.E. Wescon Convention Record, Part 6.

13. E. Keonjian; "Microminiaturization of Space Computer", the 14th Annual Meeting of American Rocket Society, Washington, D. C.; November 1960.

14. J. M. Early, "Speed, Power and Component Density in Multielement High-Speed Logic Systems"; The 1960 International Solid-State Circuit Conference, Philadelphia, Pa.; February 1960.

15. R. Norman, J. Last and I. Haas, "Solid-State Micrologic Elements"; the 1960 International Solid-State Circuit Conference; Philadelphia, Pa., February 1960.

16. F. P. Granger, Jr. and J. G. Smith, "Micro-circuitry - a Practical Technology for Reliable Microminiaturization", the 1960 I.R.E. International Convention, New York, N.Y., March 1960.

17. H. W. Henkels, "Molectronics"; The 1960 Winter general meeting of AIEE, New York, N.Y.; February 1960.

18. E. Keonjian, "Microminiature Full Adder", The 1960 Winter general meeting of AIEE, New York, N. Y.

19. D. A. McLean; "Microminiaturization with Refractory Metals", The 1959 I.R.E. WESCON Conventional Record, Part 6, pp. 87-91.

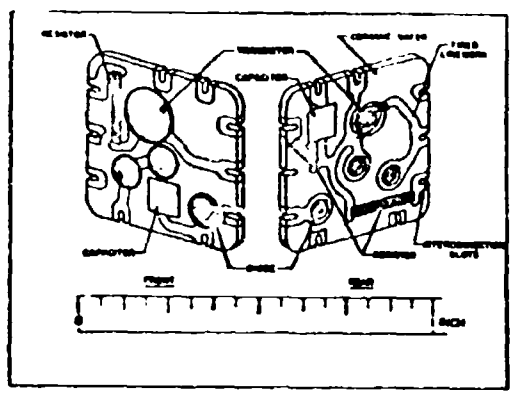


Figure 1. Single Flip Flop Wafer Layout - Carbon Composition Resistors

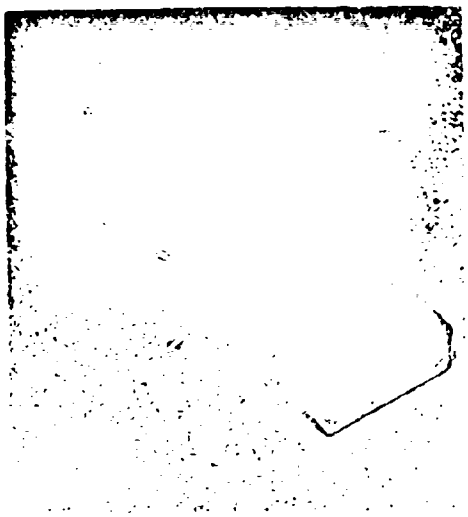


Figure 2.

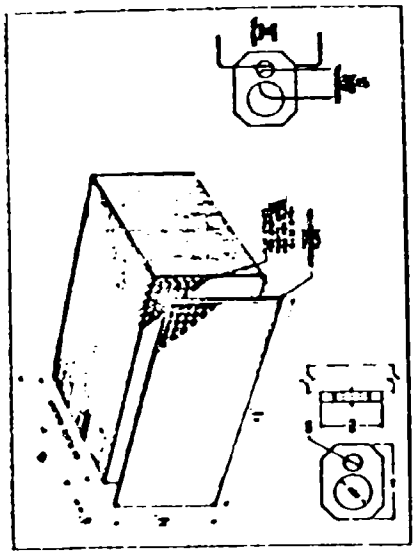


Figure 4 Diagram of Proposed Interc nnection Wiring for Reduced Wiring

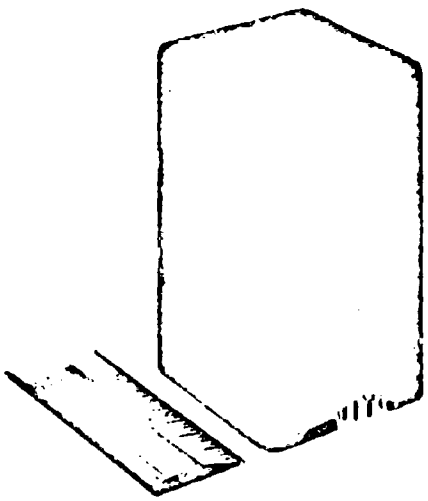


Figure 3. Typical Logic Assembly - Cutaway View

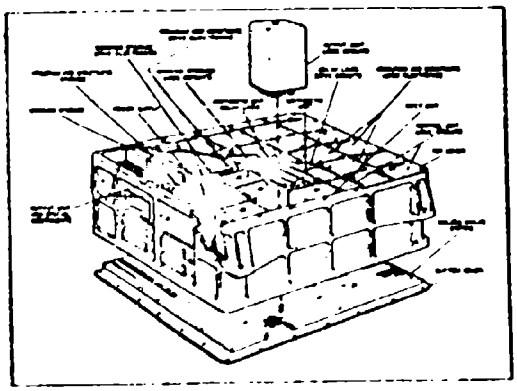


Figure 5. Exploded view of Microminiature Computer Assembly

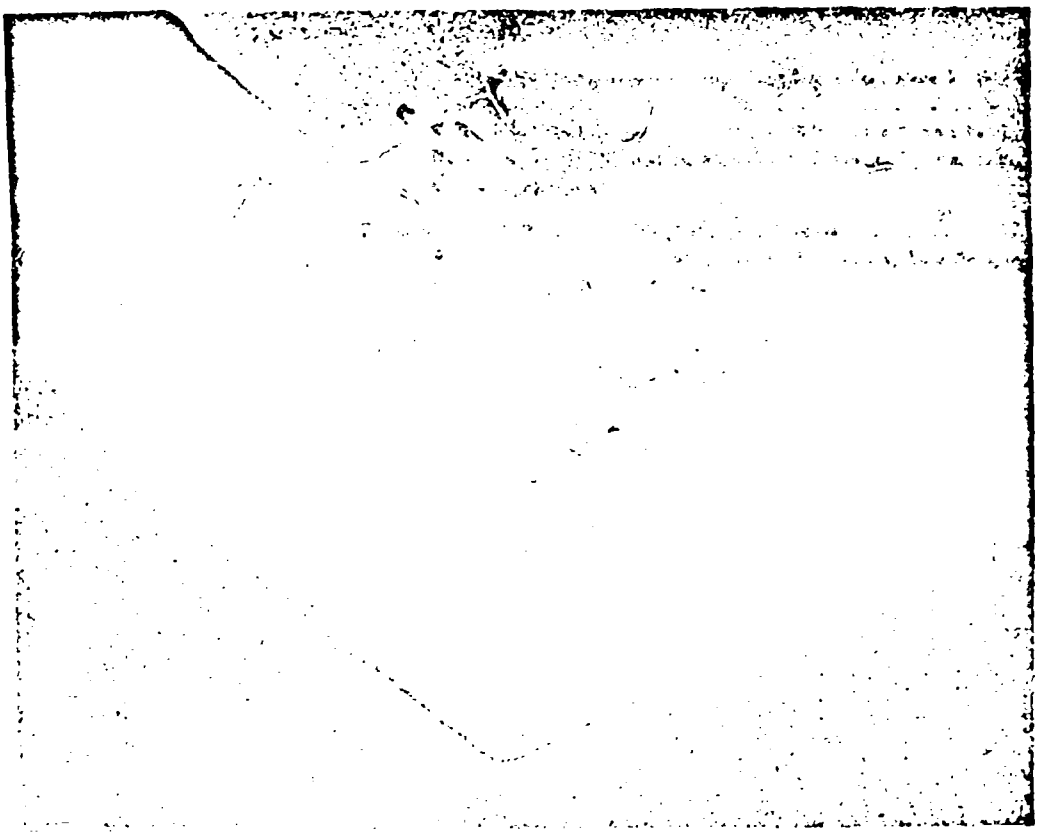


Figure 6.

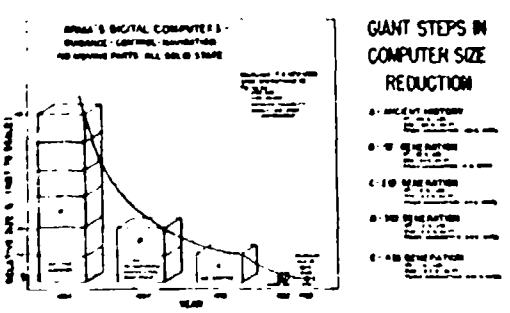


Figure 7.

CRITERIA FOR THE OPTIMUM DESIGN OF ACTIVE SATELLITE COMMUNICATION SYSTEMS

by
A. R. Giddis
Philco Corporation
Western Development Laboratories
Palo Alto, California

Introduction

This paper presents a systematic quantitative picture of major factors that must be analyzed to design optimum active satellite communication systems. Since the accent is on communications, we will not investigate the guidance and tracking of vehicles, the logistics of installations, or the economics of launching rockets.

System ground rules are established for a synchronous active relay by analyzing the following factors: (a) wave propagation and radio noise; (b) payload design; (c) performance of communication and control equipment; and (d) type of relay link. In particular, two propagation models are postulated: (1) normal weather and predicted geomagnetic activity; (2) all-weather operation during periods of severe fading. The frequency range of interest is 100 to 10,000 mc/s.

"Optimum" frequencies are derived for the following relay links: (a) satellite-to-ground links consisting of antennas that are either area-specified, beam-specified, or both; and (b) inter-satellite links consisting of antennas that are area-specified. The frequencies are selected on the basis of the maximum bandwidth that can be relayed by the satellite for specified system parameters.

The links are evaluated in terms of link capacity as a function of payload. The effects of variations and trade-offs among system parameters on the traffic-handling capacity, equipment design, and rocket booster performance are discussed.

Propagation and Noise

To begin the analysis, we will establish propagation and noise models that account for atmospheric absorption and noise, thermal radiation from the ground, galaxy, and sun, and Faraday rotation.

Atmospheric Absorption

The combined one-way absorption of the troposphere and ionosphere along an horizon path is shown in Figure 1 as a function of frequency. The lower curve is based on typical summer weather (ref. 1) and predicted geomagnetic activity for January 1964 (ref. 2). The upper curve is based on heavy rainfall (ref. 3) and severe fading typical of auroral regions (ref. 4). Between the two limiting combinations will lie the actual day-to-day atmospheric losses.

Radio Noise

Figure 2 shows the noise temperature from ground thermal radiation, tropospheric absorption, and cosmic noise at a ground antenna pointing toward the horizon.

An "average" ground (ref. 5) is assumed to radiate at an equivalent black-body temperature of 290°K into the main beam and minor lobes of a circularly polarized antenna (ref. 6).

The tropospheric noise curves are based on a narrow pencil-beam antenna with low side lobes (ref. 7) looking along a horizon path into the weather models previously mentioned.

The "average" galactic noise curve is a convenient compromise function constructed to approximate observed values (ref. 8).

Solar Radiation

The strongest natural radio source that can affect communications is the sun. Figure 3 illustrates the solar noise and the antenna noise temperature for a quiet sun under maximum sunspot conditions (ref. 9). For satellite-to-ground links it is assumed that the sun is located 10 degrees off the main beam of a 60-foot paraboloid. For inter-satellite links, the sun is assumed to radiate into -25 db side lobes; this case is illustrated by the curve for a 10-foot aperture in the synchronous satellite.

Faraday Rotation

Under normal conditions, the Faraday rotation is small in the S-band region and above (ref. 10). To protect the system against the likelihood of Faraday fading during abnormal conditions over the 100 to 10,000 mc/s spectrum, this analysis adopts circularly polarized receiving channels to detect linearly polarized transmissions. The assumed loss is 3 db over all links.

Payload Design

In addition to propagation and noise factors, the design of the payload affects significantly the performance of satellite relays. Some of the important characteristics are attitude and orbit control, configuration and size, and payload weight which ultimately depends on booster capability. The degree of attitude control determines the minimum beamwidth of satellite antennas. The configuration and size affect the shape and available area of flush-mounted and self-erecting antennas. The payload weight limits the amount of equipment and determines the power available for transmitters.

The assumed payload model is a cylinder whose topside accommodates solar cells that unfold to-

wards the sun. The base houses a wide-beam antenna that is erected or uncovered in orbit. The volume allows packaging four expandable, steerable, 10-foot antennas in addition to electronic equipment and control mechanisms. For payloads ranging from 1000 to 10,000 pounds, characteristic dimensions of the cylinder are 10 to 20-foot lengths by 5 to 10-foot diameters.

The attitude stabilization about the three axes is held to ±1 degree.

Equipment Performance

Within the framework of the payload model, the performance of communication equipment will now be established.

Vehicle Antennas

At the synchronous altitude, a ±1 degree attitude stabilization requires that the beamwidth of the antenna for horizon-to-horizon coverage be broadened from the 17.4 degrees shown in figure 4 to 19.4 degrees, in order to limit the gain degradation to 3 db along the horizon. The corresponding power gain depends on the antenna used. One useful form is a circular array transverse to the axis of the cylinder at its base. In the system analysis to follow, a gain of 19 db is adopted.

The 10-foot steerable antennas used over satellite-to-ground and intersatellite links are gain-limited by the accuracy with which they can be positioned. If the stabilized axis of the cylinder is used as reference, the ±1 degree uncertainty limits the corresponding maximum operating frequency to 3400 mc/s. If the reference is established by a direction-finding beam of 1 degree that locks onto a signal from a ground control station, an angular resolution of ±1 milliradian may be realized. The corresponding maximum frequency of operation is 60,000 mc/s.

Ground Antennas

The following assumptions are made concerning the size and gain of antennas at earth terminals: (a) over satellite-to-ground area-specified and mixed-antenna links, directional ground antennas are 60-foot paraboloids; (b) over the beam-specified links, antennas have 3 db gain in the direction of the satellite.

The minimum realizable beamwidth adopted is 1 milliradian on the basis that a beam pointing accuracy of ±1/2 milliradian represents a current limit to practical servomechanical system performance and reasonable antenna manufacturing and environmental tolerances. For typical paraboloids or linear arrays, this represents about 69 db gain. The corresponding maximum frequency for a 60-foot aperture is 20,000 mc/s.

RF Losses

Antenna system losses are assumed frequency independent and are predicated on duplexed channels and short transmission lines. A 1-db loss typifies the r-f loss at each terminal over all links and generates an excess noise of 60°K.

Receivers

The receivers are 100°K preamplifiers with a gain of 21-db and terminated by conventional 1000°K superheterodyne i-f amplifiers. Taking into account the r-f loss, the equivalent receiver noise temperature is 136°K.

Power Generation and Transmission

Primary power required to activate communication networks, attitude control and station-keeping mechanisms, and antenna beam control networks is generated by planar solar-cell arrays continuously directed toward the sun. The specific weight of this configuration is 0.15 pound of power supply weight per watt of d-c primary power, a value based on experience with solar power at MDL.

Generally, the overall efficiency, weight, and volume of transmitters decrease with increasing frequency. Exceptions to this exist and depend on power levels, the type of modulation, and the type of power amplifier (ref. 11). For example, if channel bandwidth is the predominant factor, higher frequency equipment (1000 to 10,000 mc/s) is warranted. If lifetime and stability are overriding factors, lower frequency transmitters (100 to 1000 mc/s) are preferred.

Optimum Frequencies

Applying the propagation, noise, and equipment parameters, optimum frequencies have been derived for satellite-to-ground links that consist of antennas that are either area-specified, beam-specified, or both, and for links between synchronous satellites. As shown on Figure 4, the synchronous satellites are in circular orbit at an altitude of 19,360 nautical miles. Satellite-to-ground links are assumed to exist along horizon paths. The range is 22,550 nautical miles.

The optimum frequency band is defined arbitrarily by upper and lower frequencies at which the bandwidth is degraded from its maximum by a factor of 2.

Area-Specified

Satellite-to-ground links with area-specified antennas at earth and satellite terminals are typified by the bandwidth-vs-frequency curves shown in Figure 5. Under normal conditions, the optimum frequency range lies above 10,000 mc/s but below the first water-absorption peak at 22,200 mc/s. Under abnormal conditions, the optimum frequencies lie between 2400 and 8200 mc/s. It is apparent that the gain of area-specified antennas overrides atmospheric losses until water absorption becomes severe.

Beam-Specified

Links with beamwidth (or gain) - specified antennas at both earth and satellite terminals experience relative bandwidth variations shown in Figure 6. Under normal conditions, the optimum

frequencies lie in a band from somewhat below 100 mc/s up to 300 mc/s. During abnormal weather and geomagnetic activity, the optimum band is 310 to 1100 mc/s. In this case, the decrease in space loss overcomes the increase of noise and atmospheric absorption until severe fading appears.

Mixed-Antenna

Links with a beamwidth-specified satellite antenna and an area-specified earth antenna are characterized by a constant antenna gain-space loss product, so that the optimum frequency range is determined by propagation and noise conditions. The results are displayed in Figure 7. Under normal conditions, the optimum frequencies range from 180 mc/s to somewhat over 10,000 mc/s. Under abnormal conditions, the optimum frequencies are between 600 and 5000 mc/s.

Intersatellite Links

Links between synchronous satellites consist of area-specified apertures. A typical bandwidth curve is shown in Figure 8, where no apparent upper bound exists because no atmosphere exists. Practical limitations are antenna pointing accuracy and r-f power generation.

System Evaluation

In order to calculate available transmitter power for a range of payloads, apportionment of relative power among communication networks and control systems has been made, and weight and power coefficients have been selected. These factors are a combination of known and estimated values, and are tabulated on Figure 9. From the relation shown on the figure, the slope of available transmitter power to payload is 0.108 watt per pound.

In order to operate continuously and reliably over a one-year period, reliability techniques (such as standby spares with their associated switching circuits) must be incorporated. This decreases the payload weight (and volume) available for transmitters. Assuming an average redundancy factor of three, the useful payload is reduced by 3 and the available transmitter power is 0.036 watt per pound of satellite payload.

Balanced Design

To evaluate the bandwidth capabilities of the synchronous relay satellite, we propose a relay that can support the four types of links simultaneously. An optimum relay not only should transmit a maximum bandwidth but should balance the capacity of satellite-to-ground and intersatellite links, since the latter potentially carry messages from one part of the globe to any other. Moreover, we must determine the load on the beam-specified and mixed-antenna links. If these comparatively lower capacity links are to support wide-band data transmission, a minimum percentage of the available satellite transmitter power must be assigned to these links, or similarly, a minimum payload must be orbited. For the present analysis, equal power has been apportioned to each link. In the model a pair of duplexed channels is assigned to each link to allow simultaneous transmission and reception. Therefore, one-sixth of the total power has been assigned to the beam-specified link, the mixed-antenna link, each of the two area-specified links, and each of two intersatellite links. The specific power in each channel is 6 milliwatts per pound of satellite payload.

Link Evaluations

On this basis, the bandwidth-payload curves of Figure 10 have been calculated for area-specified satellite-to-ground links at frequencies optimum for the two propagation models assumed. Representative results are as follows: a 5000-pound payload relays at a low bit error (SNR=40db) a bandwidth of 24 mc/s under normal conditions at 10,000 mc/s; and 3.8 mc/s under abnormal conditions at 5000 mc/s. Equivalent performance for intersatellite links with 10-foot antennas occurs respectively at 37,000 mc/s and 10,000 mc/s.

Figure 11 shows the capacity of beam-specified and mixed-antenna links. Under normal conditions, a 5000-pound payload relays 12 kc/s at a SNR of 20 db to non-directional stations at 100 mc/s; and 6.6 mc/s to 60-foot directional terminals at 500 mc/s. Under abnormal conditions, the beam-specified link transmits 820 c/s at 400 mc/s; the mixed-antenna link transmits 4.3 mc/s at an optimum frequency of 2000 mc/s.

Discussion of Results

The bandwidths calculated are based on the particular system models developed. Under different conditions, different values will result.

At elevation angles other than the zero degrees adopted in the analysis, atmospheric loss and noise temperature drop substantially, and ground thermal noise decreases. The range of optimum frequencies broadens with increasing elevation angle and shifts gradually to include higher frequencies.

Sky-temperature distributions, changing levels of ionospheric fading, varying weather, and location of the sun produce shifts in optimum frequencies. For a system to operate under all conditions expected, it must be designed either to meet the worst conditions or to function under reduced capability or reliability during the adverse periods.

The author has presupposed that by 1964 a 100°K receiver temperature can be realized independent of frequency in the 100 to 10,000 mc/s band. Under present technology, receiver noise figures rise with frequency. In addition to noise figures, modulation and bandwidth product are frequency weighting factors.

Overriding criteria in deciding operating frequencies today are realizable power levels and transmitter lifetime as a function of power level and duty cycle. It is expected that, since there are no inherent frequency limitations to power transmission, transmitting devices eventually will be available anywhere in the radio spectrum to generate at least moderate powers from satellites

(ref. 12). In any case, optimum power-weight and transmitter power-antenna gain relationships exist. Further study will better define these quantities, which are affected greatly by specific applications.

The apportionment of payload weight among the components of a communication system is a crucial item in the efficient design of the system. The weight and power allotted to a servo system to direct the solar array toward the sun yields a significant increase in generator power which can be used to increase available transmitter power and to supply power to standby equipment.

The degree of attitude-control affects the beamwidth of antennas designed to illuminate continuously either the whole earth visible to the satellite or portions of the earth's surface. Loose control requires broader beams and thus demands higher transmitter power or narrower bandwidths for the same signal-to-noise ratios in proportion to the square of the beamwidth ratio.

This interdependence of bandwidth, signal-to-noise, and transmitter power relates ultimately to the useful payload in orbit. A heavier payload allows in priniciple one or more of the following: (a) greater traffic-handling capacity through higher available power or more channels; (b) a greater level of reliability through redundant circuits and switching matrices and through heavier protection against the environment; (c) a longer lifetime and usefulness through reliability techniques; and (d) the feasibility of incorporating anti-jam techniques in the satellite. Of course, heavier payloads mean more powerful rockets.

Conclusion

This paper has presented the results of a systematic analysis and synthesis of major factors germane to the optimum design of an active synchronous satellite system. By considering propagation and noise, payload design, and equipment performance, optimum frequencies have been derived and used as a basis to evaluate the traffic-handling capacity of satellite-to-ground and inter-satellite links.

References

1. D. C. Hogg, "Effective Antenna Temperatures Due to Oxygen and Water Vapor in the Atmosphere", J. App. Phys., vol. 30, p. 1417; September 1959.

2. L. C. Kelley, "Long-Term Predictions of Geomagnetic Disturbances and Their Effects on Radio Communications", U.S. Army Sig. Rad. Prop. Agency Proj. No. 636A; Nov. 1958.

3. D. E. Kerr, Propagation of Short Radio Waves, M.I.T. Rad. Lab. Series, vol. 13, McGraw-Hill, p. 682 (1951).

4. C. G. Little, R. P. Merritt, E. Stiltner, and R. Cognard, "Radio Properties of Auroral Ionosphere," Qtly. Prog. Rept. No. 6, RADC Contract No. AF 30(635)-2887, Geophys. Inst. of the Univ. of Alaska, p. 5; Dec. 8, 1957.

5. H. R. Reed and C. M. Russell, Ultra-High Frequency Propagation, Wiley, N.Y. pp. 91-97;1953.

6. R. C. Hansen, "Low-Noise Antennas", The Microwave Journal, vol. 2, no. 6, pp. 19-24, June 1959.

7. R. H. Dicke, R. Beringer, R. L. Kyhl, and A. B. Vane, Physical Review, vol. 70, p. 340 (1946).

8. D. C. Hogg and W. W. Mumford, "The Effective Noise Temperature of the Sky", MARCOM Nat'l. Conf. Proc., p. 584; May 1959.

9. C. W. Allen, Astrophysical Quantities, The Athlone Press, p. 153 (1955).

10. W. A. Counter and H. F. Ridell, "Calculations of Ground-Space Propagation Effects", Lockheed Report LMSD-2461; May 22, 1958.

11. R. Strauss, "Microwave Tube Ampl'fiers for the Space Age," MARCOM Nat'l. Conf. Proc., p. 760; May 1959.

12. J. R. Pierce and C. C. Cutler, "Interplanetary Communications", Advances in Space Science, vol. I, ed. by F. I. Ordway, Academic Press, p. 66 (1959).

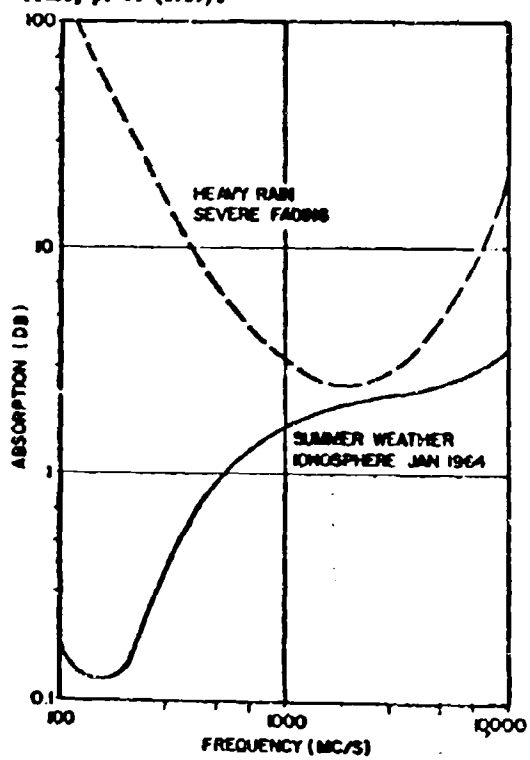


Figure 1 Atmospheric Absorption Along Horizon Path

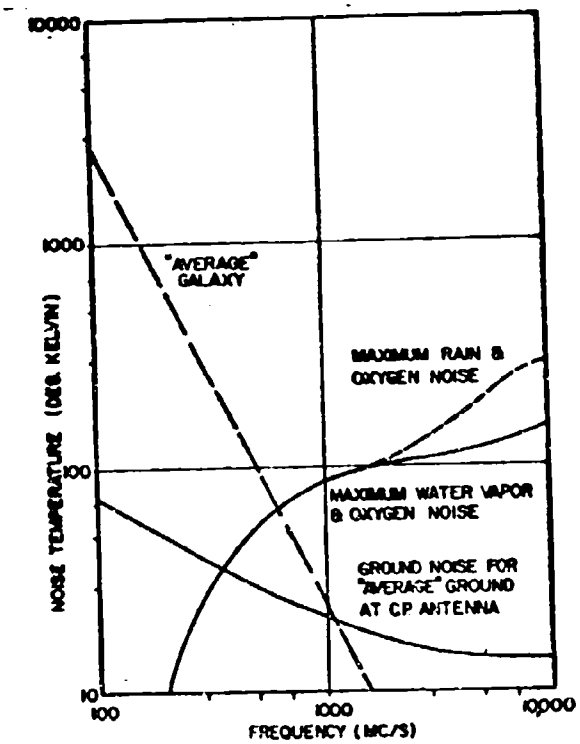


Figure 2. Noise Temperature at Ground Antenna From Ground, Troposphere, and Galaxy

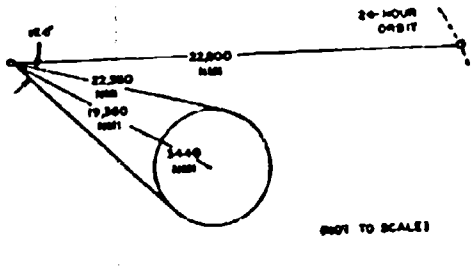


Figure 4. Satellite Relay Geometry

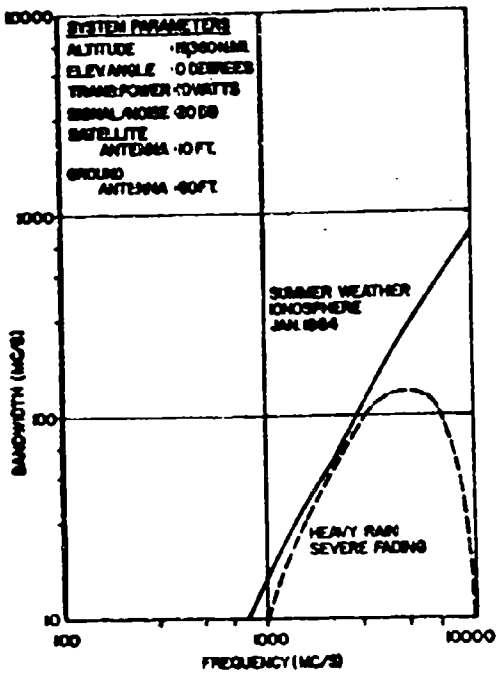
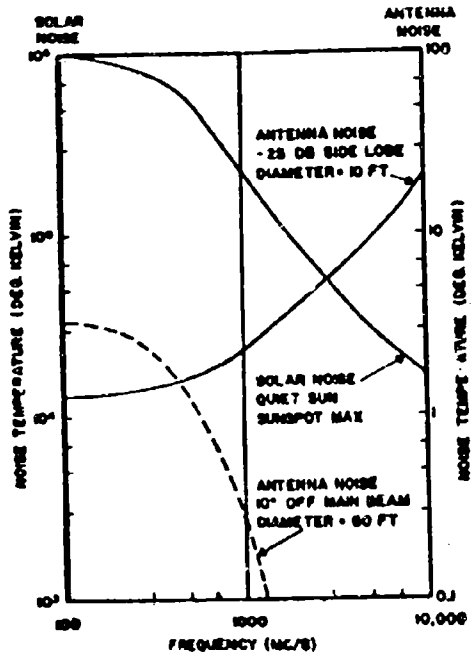


Figure 3. Noise Temperature from Quiet Sun



Bandwidth vs Frequency for Area-Specified Satellite to Ground Links

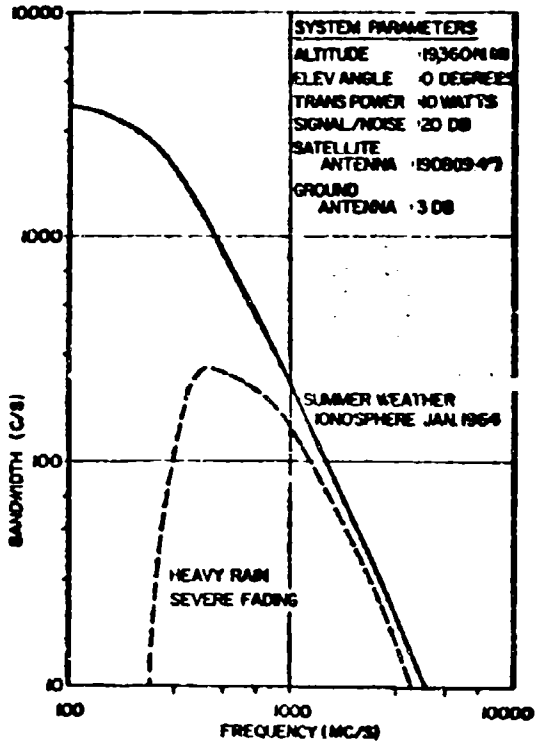


Figure 6 Bandwidth vs Frequency for Beam-Specified Satellite-to-Ground Links

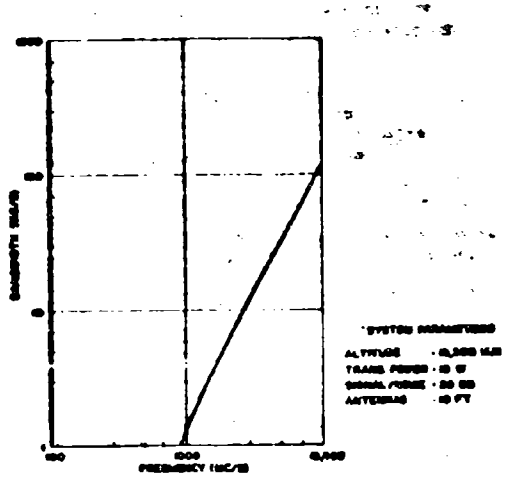


Figure 8 Bandwidth vs Frequency for Area-Specified Intersatellite Link

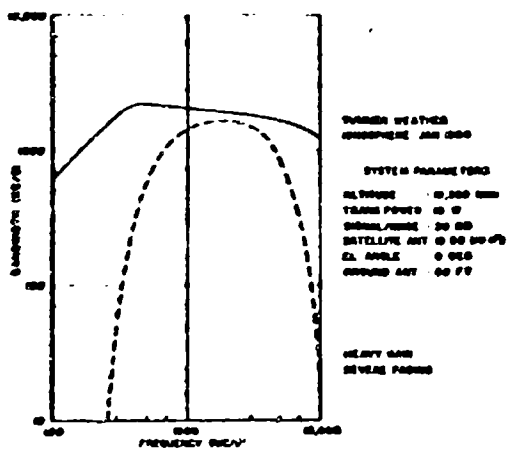


Figure 7 Bandwidth vs Frequency for Mixed-Antenna Satellite-to-Ground Links

FIGURE 9

AVAILABLE TRANSMITTER POWER

vs

PAYLOAD IN ORBIT

Conditions

Altitude: 19,360 n.mi.
Lifetime: 7 years

Weight-Power Coefficients

Power Supply: Planar Solar Arrays Directed at Sun

Efficiency, DC-RF Conversion	(d)	=	15%
Power Supply Specific Weight	(w_p)	=	0.15 lbs/watt
Transmitter Specific Weight	(w_t)	=	4 lbs/watt
Power Supply Weight/Payload	(p)	=	30%
Antenna Weight/Payload	(a)	=	5%
Receiver Weight/Payload	(r)	=	5%
Control Systems Weight/Payload	(c)	=	30%
Structures Weight/Payload	(s)	=	15%
Relative Antenna Power Drain	(P_a)	=	5%
Relative Receiver Power Drain	(P_r)	=	5%
Relative Control System Drain	(P_c)	=	5%

Formula

$$\frac{P_T}{W} = \frac{1 - (a + r + c + s)}{\left(\frac{w_p}{p} + w_t\right)} - \frac{p}{\left(\frac{w_p}{p} + w_t\right)} (P_a + P_r + P_c)$$

Results

No redundancy assumed: 0.188 watts/pound
Redundancy factor of 3: 0.036 watts/pound

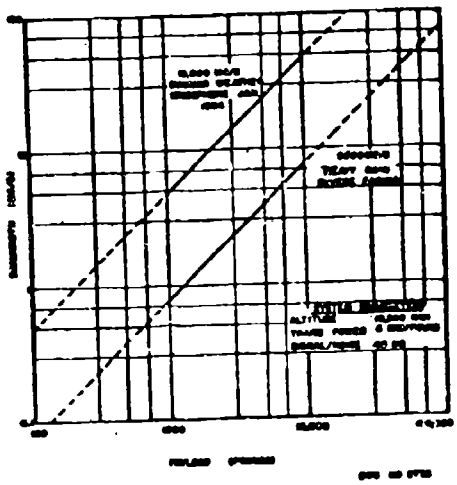


Figure 10 Bandwidth vs Payload for Area-Specified Satellite-to-Ground Links

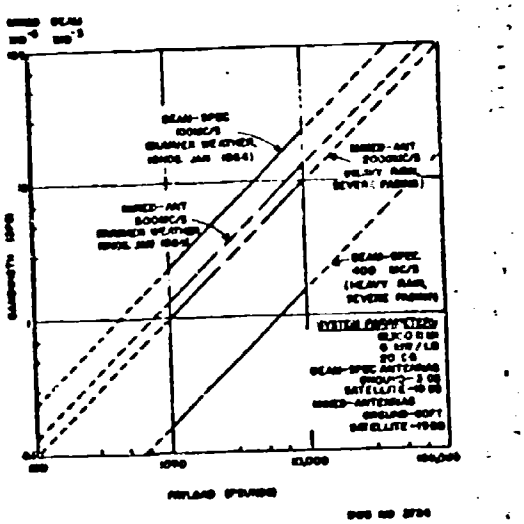


Figure 11 Bandwidth vs Payload for Beam-Specified and Mixed-Antenna Satellite-to-Ground Links

SATELLITE IONOSOUNDER

By: Samuel Horowitz, Air Force Cambridge Research Center
and Leonard Humphrey, General Electric Company

1. INTRODUCTION

Knowledge of the electron density profile above the F-layer maximum is of extreme interest both for its application to space communication problems, and for the extension of our fundamental knowledge of the origin and structure of the earth's ionospheric envelope. Clues to the origin of such phenomena as spread-F, sporadic E, the Van Allen belt or aurora may well be hidden in the upper ionosphere. Also, in this age of space communication and navigation, a knowledge of ionospheric electron density at all altitudes is essential for the determination of refraction and multipath effects, polarization rotation, severity of vehicle induced plasma sheaths, antenna perturbations, and other similar considerations.

The determination of this high altitude electron density from the ground is possible using electron backscatter techniques as suggested by Gordon of Cornell University. The required power is high and antenna sizes are large so that such devices will be limited to a few geographic locations. The use of conventional ionospheric sounding techniques from a high altitude polar orbit satellite will permit obtaining knowledge of the upper ionosphere over the entire earth. There are many degrees of sophistication that such a sounder may possess, the design of each further sophisticated type sounder being influenced by the results obtained by earlier types. Early type sounder would be rockets or satellites continuously telemetering data to give a complete picture of the ionospheric profile in the vicinity of the telemetry receiver as is presently being accomplished by the Propagation Laboratory at AFCRC. The next order of sophistication would be a continuously recording sounder which would telemeter the results obtained during its entire orbit, thereby extending our knowledge geographically. The final order of sophistication would be specialized types for ICBM and nuclear detonation detection or recording desired parameters such as muf.

The satellite ionosounder discussed here is of the second order sophistication. The information recorded will be "A" scan type so that the magnitude of the returns as well as the time delay will be determined. The information will be recorded over the entire polar orbit so that eventually a complete latitude-time dependent electron density profile will be obtained.

2. BASIC DESIGN PHILOSOPHY AND PRELIMINARY CONSIDERATIONS

The prime mission of the sounder is to gather information about the electron density distribution of the upper ionosphere. The design of the sounder must be such that sufficient data to accomplish this mission can be reliably collected and returned to ground. For reliability of the sounder equipment itself, simplicity of design with a minimum of vulnerable components; for reliability of the vehicle and booster a minimum weight payload with minimum stabilization and orbit requirements is planned.

2.1 Preliminary Orbit Considerations

To acquire information over the most widespread geographic area, a polar orbit is essential, since it alone permits sounding over the entire latitude range. A polar orbit also facilitates interrogation on every pass by high latitude ground stations.

In the interest of simplicity and reliability of launch, a near circular orbit should be attempted but not required. The prime altitude requirement should be a perigee sufficiently high for a long satellite lifetime. Other factors bearing on the height of the orbit will be discussed later.

2.2 Stabilization Considerations

A dipole sounder antenna is anticipated both for simplicity and in order to excite both modes of propagation. Maximum signal return occurs when this dipole is oriented parallel with the reflecting plane. A stabilized platform with horizon sensors could be used to keep the antenna properly oriented, but the weight and power requirements of such a system are prohibitive.

Spin stabilization is much more economical in weight and power, and with proper design can be almost as effective as the sensor stabilized system. If the antenna rotates about an axis perpendicular to the antenna itself, it may be seen that at least twice in every rotation it is parallel to any arbitrarily oriented plane which would represent the earth's ionosphere. The vehicle will rotate stably about the axis of maximum moment of inertia, and if the antenna is mounted perpendicular, the desired stabilization will be achieved.

Summarizing, the following discussion will have as a starting point a spin stabilized vehicle

in a near circular polar orbit. More comprehensive considerations are required to determine more detailed specifications.

3. ORBIT PARAMETERS

It has been determined above that the satellite orbit should be polar and near circular for reasons of simplicity and maximum data collection capability. The remaining orbit parameters to be determined is the orbit radius or satellite height. Figures 1 through 5 show graphically the effect of satellite height on system requirements.

Figure 1 shows the minimum visibility time during which the information may be telemetered to ground stations at various latitudes. At Anchorage, for example, which is at about 60° latitude, the minimum interrogation time (transverse orbit) is about 1/2 hour for a 3500 Km high satellite. Figure 2 shows the fraction of the time the satellite is above the horizon. This fraction is amount of time compression needed on play back and is equal to the reciprocal of the bandwidth increase factor. With the previous example this compression is about 0.2 so that the bandwidth recorded is increased by a factor of five when it is real bnck-: telemetered. The last consideration to orbit height effects on the recorder is the required storage time. Figure 3 shows the storage time required as a function of height. For a 3500 Km circular orbit and a 60° latitude interrogation station about 140 minutes of storage is required.

The telemetry power requirements depend on the maximum slant range to the satellite during the interrogation period. Figure 4 shows this range as a function of satellite altitude.

The sounder pulse power requirement is also a function of satellite height. Figure 5, which is discussed later in more detail, shows the pulse power required for satisfactory sounding.

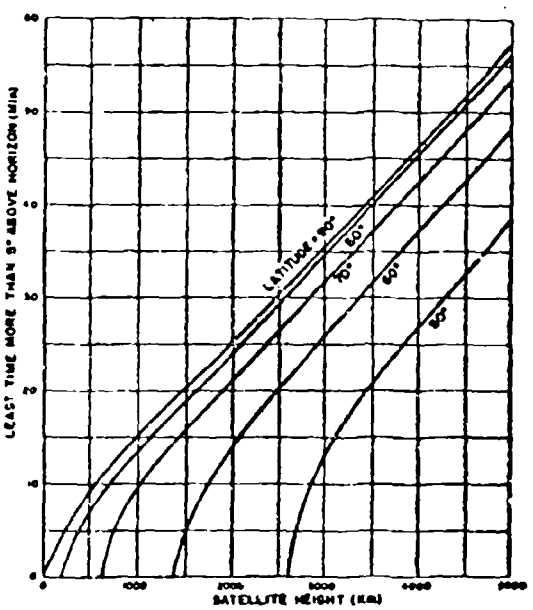


Fig. 1 - Least Time More Than 5° Above Horizon (Interrogation Time) vs Satellite Height

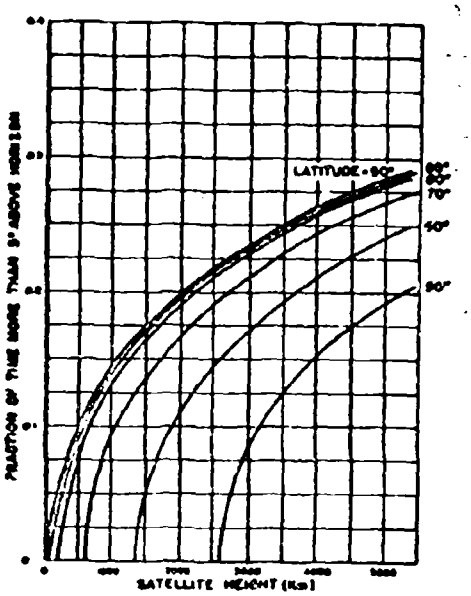


Fig. 2 - Maximum Time Less Than 5° Above Horizon vs Satellite Height

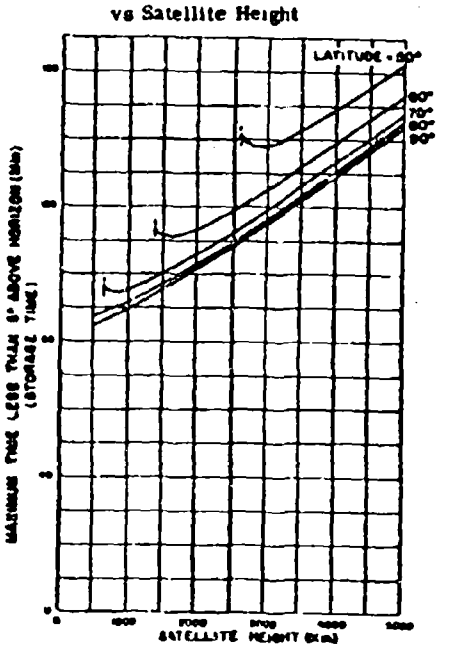


Fig. 3 - Fraction of Time More Than 5° Above Horizon vs Satellite Height

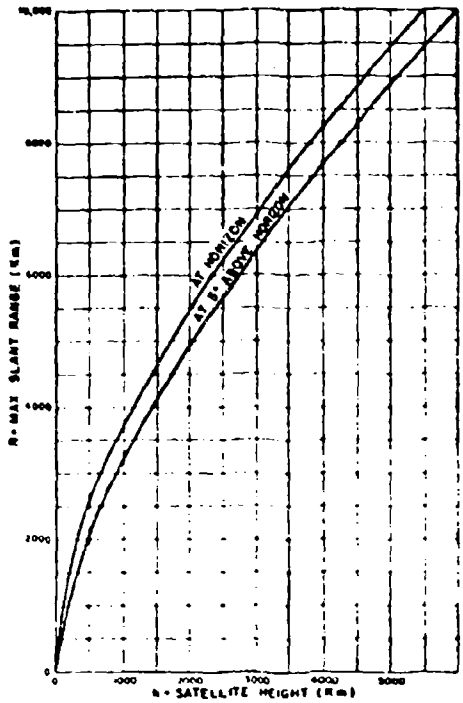


Fig. 4 - Maximum Slant Range
vs
Satellite Height

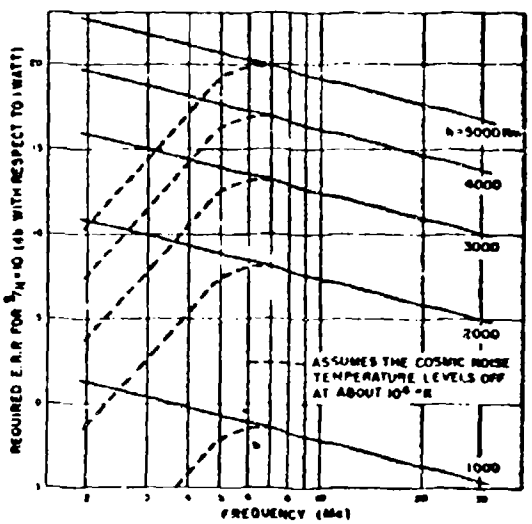


Fig. 5 - Power Requirement for S/N = 10;
10 kc Bandwidth

From these curves, it is seen that higher heights are desirable from a storage and playback viewpoint, but that lower heights are desirable from the viewpoint of the sounder itself. For high altitudes, telemetry power requirements are higher, but bandwidth requirements are less. The problem of deciding on the best height reduces to selecting that height for which reliable, lightweight components (storage, telemetry or sounder) exist.

In addition to the considerations given in these curves, considerations such as orbit lifetime, antenna detuning, or effect of radiation belts should be taken into account.

4. PROPAGATION CONSIDERATIONS

4.1 Assumed Electron Density Profile

There is data available from rocket ionprobes, Faraday rotation from satellites, whistler measurements, and electron backscatter experiments from which some idea of the distribution of electron density above the F2 maximum can be deduced. Figure 6 shows an electron density curve based on these data. It must be realized that this distribution can vary markedly in shape and magnitude with time and geographical position. This distribution can be used, however, to get an idea of the h' - f records that will be obtained with the sounder.

4.2 Expected h' - f Profile

In the absence of magnetic fields, the phase velocity of a radio wave is given by

$$v_\phi = \frac{C}{\sqrt{1 - \frac{f_N^2}{f^2}}} \quad \text{where } f_{N(KC)} = 9\sqrt{N} \quad \text{and } N = \text{electrons/c.c.} \qquad (1)$$

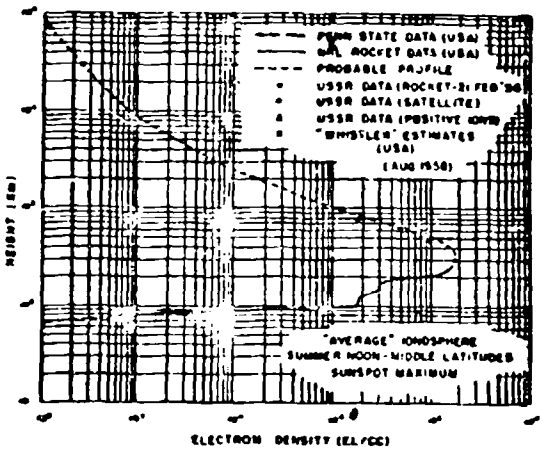


Fig. 6 - Electron Density Profile

With no magnetic field present, the group velocity is

$$v_g = \frac{c^2}{v_\phi} = c\sqrt{1 - \frac{f_N^2}{f^2}} \qquad (2)$$

The wave is reflected when N is large enough that $f_N = f$. Since N is a function of h, the reflection height is a function of f, the operating frequency. If we designate this height h_R, the round trip pulse delay is given by

$$\Delta T = 2\int_{h_R}^{h_s} v_g \, dh = 2C\int_{h_R}^{h_s}\sqrt{1 - \frac{f_N^2}{f^2}}$$

$$dh = 2C\int_{h_R}^{h_s}\sqrt{1 - \frac{81N}{f^2}}\, dh \qquad (3)$$

The equivalent height is defined as the height at which a pulse travelling at the speed of light would be reflected and in the case of ground sounder ($h_s = 0$) would be $h' = \Delta T/2C$. It is convenient to use the same form for the sounder, with the understanding that equivalent height, h' is actually the equivalent depth of reflection below the sounder, i.e. :

$$h' = \frac{\Delta T}{2C} = \int_{h_R}^{h_s}\sqrt{1 - \frac{81N}{f^2}}\, dh \qquad (4)$$

A plot of h' as a function of f is the data that will be obtained by the sounder.

In the presence of the earth's magnetic field the transmitted wave is generally split into two modes, called the ordinary and extraordinary modes. Except in the region of the geomagnetic equator, these modes are nearly circularly polarized so that each mode will be excited by a linear dipole. To a first approximation the h' - f record of the ordinary mode is the same as that of the h' - f record of the no magnetic field case. The approximate h' - f record of the extraordinary mode is given by

$$h' = \frac{\Delta T}{2C} \int_{h_R}^{h_s}\sqrt{1 - \frac{81N}{f^2(1 - \frac{f_H}{f})}}\, dh \qquad (5)$$

where f_H = gyromagnetic frequency

The (h' - f) records for h_s = 4000 Km derived from the electron density curve of Figure 6 and the approximations of equations (4) and (5) are shown on Figure 7. Since the gyromagnetic frequency is a function of the magnetic field, the calculations become laborious as the satellite position changes. There are two noticeable differences from the usual h' - f records obtained from ground sounders - the strongly decreasing h' at low frequencies and the double value characteristic of h' at intermediate values. Both these effects are attributable to the extended region of ionization in the top side as opposed to the sharply increasing ionization on the bottom side.

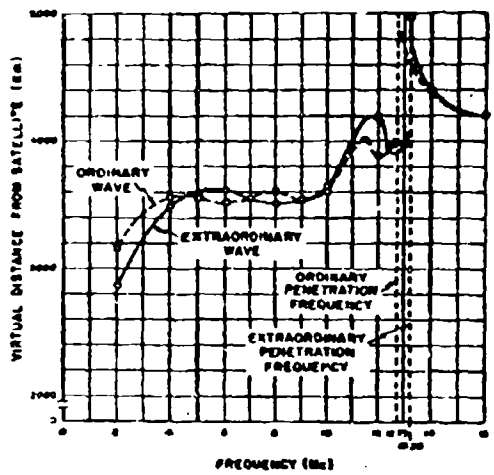


Fig. 7 - Anticipated Ionogram Satellite at 4000 Km

4.3 Application to System Design

Although the data and equations used to drive the curves of Figure 7 are approximate and subject to variation, a number of conclusions applicable to system design can be drawn from these curves:

a. Faraday Rotation. Since Faraday Rotation depends on interference between ordinary and extraordinary curves, there will be no Faraday Rotation when the time difference of arrival between these modes is greater than the pulse width. For nominal pulse width of 100 microseconds, it may be seen that there is little problem in rotation of the plane of polarization.

b. Loss due to Magneto Ionic Splitting. This is a corollary of the property first discussed above. Since the modes are propagated separately, and since the modes are circularly polarized a loss of 3 db will be experienced between a given mode and the antenna both on transmission and reception.

c. Additional Data for True N Determination. The additional data contained in the additional mode can be used to more accurately determine the electron distribution when the complete h' - f curve is not available. It is desirable, then to excite both modes as is done with a linear dipole.

d. Spacing of Sample Frequencies. The evidence of large curvatures at both the low and high frequencies with fine structure in between suggest that a sweep frequency sounder be used or if a stepped frequency sounder is used that the spacing between frequencies be small and uniform.

e. Repetition Rate and Pulse Widths. Bounds on the repetition rate and pulse width are determined by the maximum and minimum return times respectively. The actual values may fall within these bounds as determined by other considerations.

f. Antenna Detuning. Since the phase velocity of radio waves is different in an ionized medium than in free space, an antenna cut to resonate in free space may not resonate in the ionosphere. Figure 8 shows the amount of detuning expected for various heights as derived from the electron density curve of Figure 6. The model ionosphere is based on composite data from Bowles and Townsend.

5. PULSE POWER REQUIREMENTS

The pulse power required was shown in Figure 5. This power was computed on the basis of a 10 to 1 signal to cosmic noise ratio being sufficient for satisfactory analysis of the sounder records. The expected cosmic noise is shown in Figure 9. At frequencies greater than 10 mc, the emperical shape of the cosmic noise temperature curve is $T = 2 \times 10^4 (\lambda/10)^{2.5}$. Below 10 mc, the cosmic noise temperature is difficult to determine due to ionospheric screening. Figure 9 shows two schools of thought on lower frequency noise - the continuation of the emperical formula, and a cutoff at 10^6° K. Rocket noise measurement experiments being conducted by AFCRC should shed light on the low frequency behavior.

The transmission loss of the transmitted pulse power is given by:

$$L_t = L_{fs} + L_D + L_{pol} \qquad (6)$$

Where:

L_{fs} = free space loss = $-20 \log \lambda + 20 \log r + 22$ db

λ = wavelength

r = total path length = 2 h

L_D = divergence loss = $20 \log_{10}(\frac{h}{a} + 1)$

h = satellite height

L_{pol} = loss due to linear antenna coupling to circularly polarized mode = 6 db total (3 db transmit, 3 db receive).

To find the required pulse power for S/N = 10

$$P_T(\text{dbw}) = P_N(\text{dbw}) + L_t + 10 \qquad (7)$$

where P_N = noise power from figure 8.

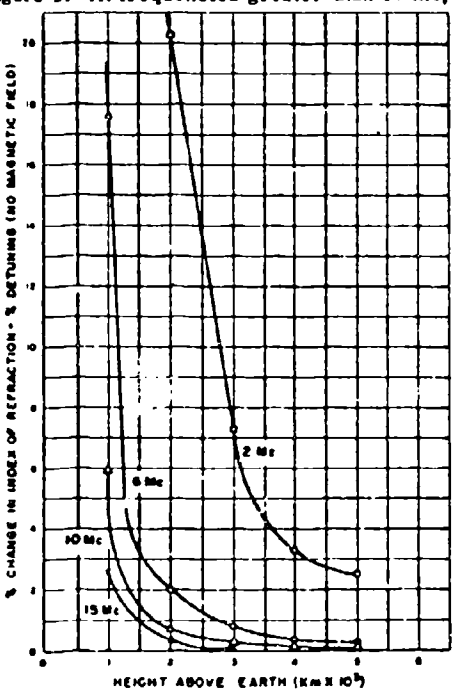


Fig. 8 - Antenna Detuning

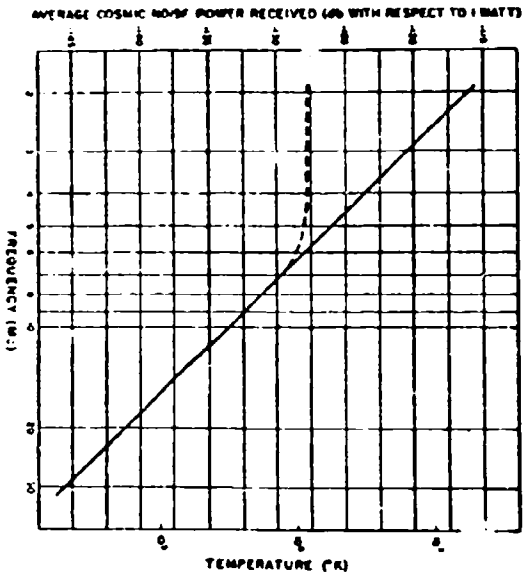


Fig. 9 - Cosmic Noise for 10 KC Bandwith

Figure 5 is a plot of equation 7 for various altitudes. We have assumed that there is no ionospheric absorption since collision frequencies are small above the F layer. We have also assumed that there is no loss on reflection other than the divergence due to the convex surface. Local irregularities may cause additional loss, but on the average this loss is expected to be low.

Since the pulse power discussed above involved radiated power only, antenna gain and circuit loss are not included. On transmission, the circuit loss is expected to balance the dipole gain so that the transmitter output power of the sounder should be approximately the same as the radiated power. Since the cosmic noise temperature is so high, the circuit and receiver noise is considered negligible.

6. SYSTEM SPECIFICATIONS

6.1 Pulsewidth

As a starting point, a pulse width of about 100 microseconds appears to be optimum. The factors considered in arriving at this figure are desired resolution, power requirements, receiver bandwidth and pulse dispersion.

a. Resolution. A short pulsewidth is desirable for good resolution. A 100 usec pulse gives about 10 mile resolution that is considered adequate for this application.
b. Power Requirements. Long pulses are desirable for easing peak power requirements, since long pulses allow the use of smaller receiver bandwidth. Since it is not desirable to compromise the resolution discussed above, a 100 microsecond pulsewidth with 10 kc bandwidth was assumed in the computation of required pulse power.
c. Receiver Bandwidth. Another limitation pulse width is receiver bandwidth and oscillator stability. To take advantage of the narrow bandwidth of long pulses, the receiver bandwidth must be small, and oscillators must be stable enough that the signal always falls in the receiver passband. Here again 10 kc is a reasonable limit. It is appropriate to mention here that doppler shift is not expected to be a problem at these frequencies.
d. Pulse Dispersion. If the h' - f record has steep-sloped characteristics, the wideband frequencies of a narrow pulse may be returned at different times, thereby broadening and distorting the pulse. By examining Figure 7, the dispersion of 10 kc wide signals is small over most of the frequency range.

6.2 Orbit Height

The philosophy in picking the orbit height was to use as high an orbit as possible to ease telemetry, storage, antenna tuning, and guidance requirements. There are currently available transistors capable of 100 watts output in the HF range. Referring to Figure 5, it appears that a 4000 Km orbit will be within the requirements of a transistorized final amplifier.

6.3 Repetition Rate

From Figure 7, the pulse delay expected for a 4000 Km high sounder is about 40 mc. However, near the critical frequency much longer delays may be expected. To acquire data in these regions, a 100 millisecond period or 10 cps repetition rate is desirable.

6.4 Storage Requirements

If we allow 5° leeway above the horizon for initiation of interrogation and playback, the recording time will be as shown on Figure 3. For logistic reasons, it may be desirable to use an interrogation station as low as 60° in latitude. For such a configuration 140 minutes recording time is required. The playback time is given in Figure 1 as 37 minutes or less. Allowing some margin, the bandwidth compression will be about 5 to 1, so that on playback a 50 kc signal can be expected.

6.5 Telemetry

The maximum bandwidth required by the telemetry channel is 100 kc assuming double sideband, AM, or narrow band FM, telemetry. For this bandwidth, and the slant range of 8000 Km given by Figure 4, the telemetry transmitter power output of 3 watts at 100 mc should be more than adequate (20 db S/N) if an 18 db ground receiving antenna is used.

For acquisition and tracking it is desirable to have a continuously operating telemetry channel. It is convenient to send back the real time data on this channel (1 watt for 20 kc bandwidth) so that a record of the ionosphere may be obtained through this channel at the same time the stored data is read out. This channel also increases reliability since data may still be obtained in the event of recorder failure.

The command receiver may be lumped with telemetry. This poses no particular problem since only narrow band information is expected to be received. The command receiver will operate continuously to be ready for interrogation from the ground station.

6.6 Power Supply

Although high pulse powers are used, the duty cycle is low so that the total power required for

Table I System Specifications	
Vehicle:	Spin Stabilized - Sounder antenna perpendicular to spin axis
Orbit:	4000 Km near circular polar orbit
Sounder:	
1. Pulse width	100 microseconds
2. Repetition rate	10 cps
3. Peak power	100 watts
4. Receiver bandwidth	10 kc
5. Total required power	3 watts
6. Antenna	Tuned dipole
Recorder:	
1. Storage	140 minutes
2. Read out	30 minutes
3. Required power	4 watts
Telemetry:	2 channels
1. Channel 1 video B.W.	50 kc
2. Channel 1 power	3 watts
3. Channel 2 video B.W.	10 kc
4. Channel 2 power	1 watt
5. Total required power	6 watts
Programmer-Command	
1. Power required	2 watts

the sounder should not exceed 3 watts. The telemetry transmitted power will be of the order of 6 watts - 2 watts continuous and 4 watts additional on readout. The recorder is expected to take 4 watts and the programs with associated circuitry, 2 watts. The total power requirement is then 15 watts. Since a power ratio of 7.5 watts/sq foot is available from solar batteries, solar power is feasible, even for tumbling satellites. The weight for a non-oriented solar supply of 15 watts should not exceed 50 pounds.

The general specifications are summarized in Table I.

7. SYSTEM DESIGN

Figure 10 shows the overall block diagram as well as a more detailed block diagram of the sounder itself. The programmer is essentially a clock and gate forming circuit as well as a controller. The sounder pulses are formed by gating the mixer which mixes a stable 18 mc signal with a signal displaced by the desired frequency. This second signal is also used as the local oscillator of the receiver so that the receiver IF is also 18 mc. By using a fairly high IF, image suppression is simplified so that S/N is not degraded. However, crystal IF filters are required to attain the 10 kc bandwidth at these frequencies.

Crystal oscillators are used for both the 18 mc and the offset oscillator due to the tight frequency requirements. Since crystal oscillators require considerable time to start and stop, the mixer is gated with the 100 usec pulse. The 18 mc oscillator must be turned off after each pulse so that it doesn't feed through the IF. The 16 mc oscillator continues to run while a number of 2 mc pulses are transmitted and provide the local oscillator frequency for the receiver. The programmer keys on the 15 mc oscillator when 3 mc pulses are required, etc.

Logarithmic IF and video amplifiers are used to prevent overload and to compress the signal dynamic range to match the dynamic range of the storage. This compression will be delayed so that S/N is not degraded.

The other components in the block diagram are conventional components, modified where necessary to satisfy the specifications in Table I. Using proven components plus the use of parallel components in the sounder circuits should assure high reliability.

The components are distributed in a disc or annular ring shape, the axis of which is the axis of maximum moment of inertia about which the satellite will spin. A long HF antenna is wrapped around the outside and allowed to unwind after vehicle separation. The stable configuration is that yielding the maximum moment of inertia

which is that of the extended antenna. Traps are inserted to tune the antenna at all frequencies.

The experimental model of the satellite ionosounder is scheduled for test at the end of this year.

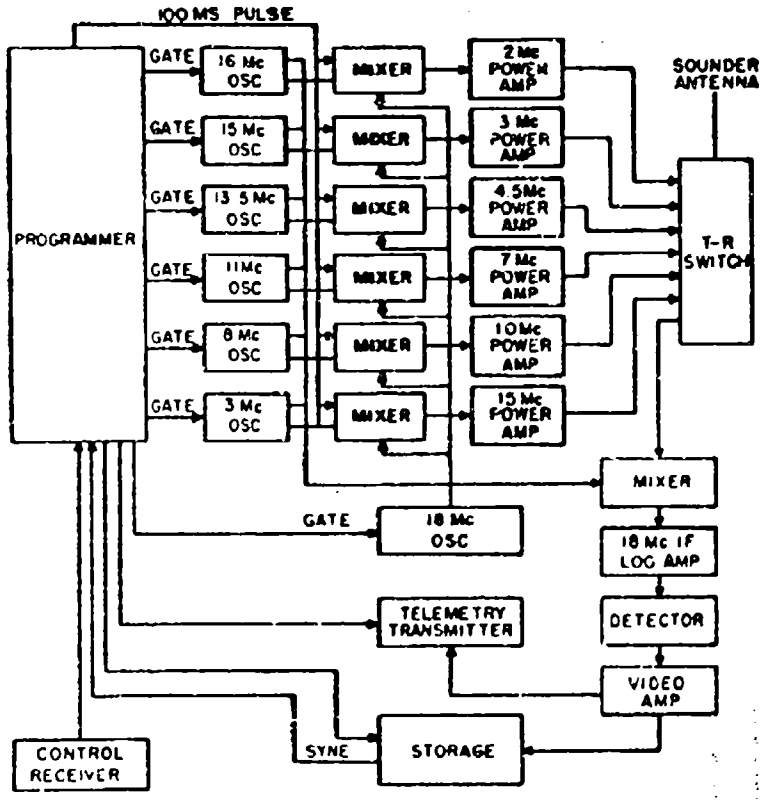


Fig. 10 - Sounder Electronics Block Diagram

PHASE STABLE LIMITING I-F AMPLIFIERS USING BEAM DEFLECTION TUBES

By: Dr. E.R. Wingrove, Jr., Electronics Laboratory, General Electric Company

Introduction

In many radar and communication systems, information is carried by the signal phase. The received signals pass through limiting i-f amplifiers which amplify the signals of varying amplitudes to a uniform level. If the phase information is not to be distorted, the limiting action must not alter the phase relationships between the various signals or signal components. Therefore, the phase shift of the limiting amplifiers must not vary with input amplitude. A limiting amplifier having a phase shift independent of input signal amplitude is defined as a phase stable limiting amplifier.

Circuits that are commonly used as limiting amplifiers inherently produce phase variations which prevent these circuits being used in systems requiring precise phase measurement. By using beam deflection tubes, however, limiting amplifiers that are highly phase stable over large ranges of input amplitude can be built.

Causes of Phase Instability

A limiting amplifier makes use of the non-linear transfer (output vs input) characteristic of a limiting device. Ideally, this characteristic would be two straight lines; a linear region, and a limiting region where output is constant, independent of input (Figure 1). The point of transition between these two regions is called the threshold. For an actual limiting device, the transition is rounded and the output is not entirely independent of input in the limiting region. However, excellent limiting amplifiers can be built with non-ideal limiters.

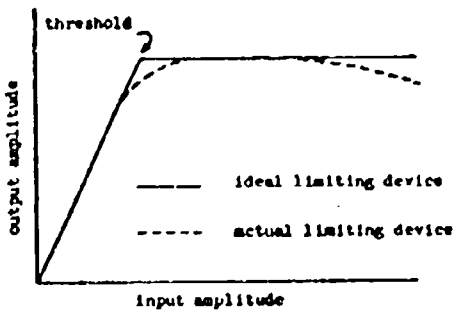


FIGURE 1 TRANSFER CHARACTERISTIC OF LIMITING DEVICES

The range of input signals over which the output of a limiting device is reasonably constant is called its limiting range. A typical limiting range is 20 db. In many systems, however, a limiting amplifier is required to amplify to the same level signals that may vary over a range of 80 db or more. To achieve such large limiting ranges, a number of limiting devices are connected in cascaded stages. In this way, the weakest signals are limited by the last stage only, the other stages acting as linear amplifiers. For stronger signals, additional stages limit, and for the largest signals, all stages limit. Limiting devices can be cascaded directly if they have sufficient gain in the linear region. If the gain is low, linear amplifier stages are necessary between limiting stages.

Band-pass networks, usually single-tuned or double-tuned circuits, couple the cascaded stages of a limiting i-f amplifier. The input and output impedances of the limiting device necessarily add to the coupling circuit impedances, and in an amplifier designed for a high gain-bandwidth product, are significant parts of the tuned circuits. If the input or output impedances change, the response of the coupling circuit will change and the phase shift through the circuit will also change. The input and output impedances of the limiting device used in a phase stable limiting amplifier, therefore, must not change with signal amplitude.

Any limiting device which limits because its input impedance changes with input amplitude is, therefore, obviously unsuitable for phase stable limiting. Devices in this category are diode clippers and grid limiters. These depend on a change in impedance to load the signal source and thus prevent the amplitude from rising above a certain level.

Other devices commonly used in limiting amplifiers produce a limiting action because of the limitation of the available output power. Examples are an overdriven transistor stage, overdriven pentode amplifier, and the gated beam tube. Phase stability and limiting could be achieved by using such a saturating device with very low impedance coupling networks. Variations in limiting device impedance would thus be "swamped" by low circuit impedance. However, the impedance changes that occur at both the semiconductor junction and vacuum tube grid are, unfortunately, so large that gain-bandwidth product would be reduced by a factor of 100 or more if circuit impedances were made low enough to ensure a phase stability of one degree or less. Impedance changes of the limiting device cause phase instability and can only be eliminated by seriously degrading circuit performance.

The second serious source of phase instability in a limiting amplifier is stray coupling of the input signal around the limiting device to its output. The most common form of this stray coupling is capacitance between input and output, both in the limiting device itself and in the associated wiring. Figure 2 shows how stray capacitance causes phase instability. The

output of the limiting device is the sum of the desired signal and a quadrature signal coupled through the stray capacitance. The direct signal, having been limited, has constant amplitude, whereas the stray signal amplitude changes in proportion to the input amplitude. Thus the phase of the resultant output varies with signal input amplitude.

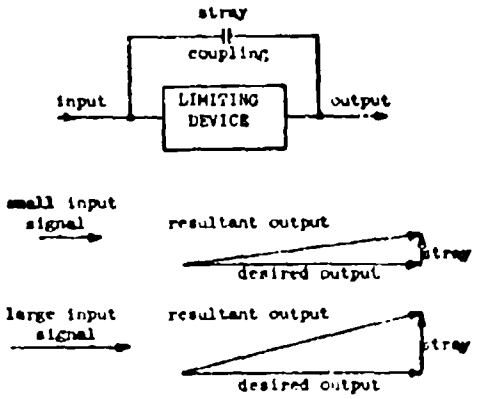


FIGURE 2 PHASE INSTABILITY DUE TO STRAY COUPLING

Unlike the input and output impedance changes, this stray coupling effect cannot be made negligible by the addition of other impedances. It can only be reduced (by careful shielding both in the limiting device and in the surrounding circuitry), or cancelled (by introducing an equal and opposite stray coupling). The first approach is desired, since the cancellation scheme adds to circuit complexity and reduces the gain-bandwidth product.

The Beam Deflection Tube

The requirements of a limiting device for use in a phase stable limiting amplifier can be summarized as follows:

a) The device transfer characteristic must have a linear region and a limiting region. Preferably gain is high in the linear region so that stages may be cascaded without need for intermediate amplifiers.

b) Input and output impedances must not vary significantly with input signal amplitude.

c) Stray coupling of input to output must be negligible.

The beam deflection tube satisfies these requirements perfectly.

In a beam deflection tube, electrons emitted from the cathode are focused into a sheet beam by an electron lens system and directed between deflection plates to a pair of anodes. A simplified sketch of the interior of a beam deflection tube is shown in Figure 3.

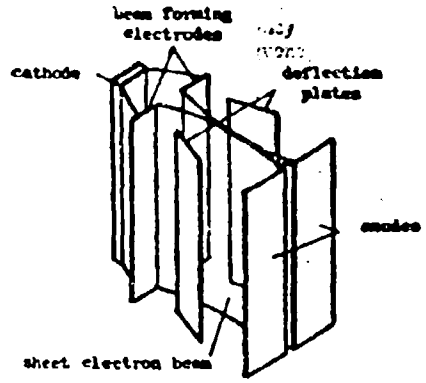


FIGURE 3 SIMPLIFIED SKETCH OF A SHEET BEAM DEFLECTION TUBE

A balanced input signal is applied across the deflectors and the output is proportional to the difference between anode currents. When the input signal is zero, there is no deflection, and the beam is divided equally between the two anodes. The output is then zero. A small input signal will deflect the beam only partially from one anode to the other and an output proportional to input will result. The tube is then in its linear region and operates as a linear amplifier.

For sufficiently large input signals, the beam is deflected completely from one anode to the other. Regardless of the peak deflection, each anode current changes from zero to the full beam current. The tube is now in its limiting region, and output amplitude is constant, independent of input amplitude.

The largest signals that the tube can handle are those which deflect the beam so far that it begins to strike the deflection plates. Part of the beam current then goes to the deflection circuit, causing a drop in output amplitude and an increase in input conductance. This point of overdriving and the threshold point (where the beam swings completely from one anode to the other) define the limiting range of the beam deflection tube.

There should, theoretically, be no capacitance change with input signal at the deflectors or the anodes, and measurements verify this theory. Capacitance changes, if any, are less than the resolution of the measurement equipment, or less than .05 pf.

Input conductance is negligible provided the tube is not overdriven. Output conductance does decrease as the tube is driven from its linear region into the limiting region, but this change is negligible compared to circuit conductance.

A shield grid, inserted between the deflectors and the anodes, does not affect the operation of the tube but reduces internal capacitance between the deflectors and the anodes. Assuming reasonable care in circuit construction, stray coupling between input and output is low enough to allow good phase stability.

Practical Circuits

The General Electric Company Electronics Laboratory has built phase stable limiting amplifiers using experimental beam deflection tubes. A small quantity of these tubes (designed by Dr. R. Adler) were prepared for GE by Zenith Radio Corporation. This tube (type 1095) has a conventional nine-pin miniature envelope. It has a gain-bandwidth product of over 40 mcps which makes possible direct cascading of stages for bandwidths up to about 15 mcs.

The important features of the circuit used with the 1095 are shown in Figure 4, a simplified schematic diagram of a single limiting amplifier stage. The symmetrical nature of the tube makes use of balanced circuitry appropriate. An advantage of balanced circuitry is that the effect of the electrode capacitances of the tube are nearly halved.

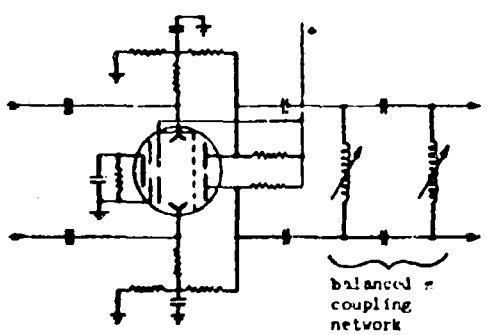


FIGURE 4 SIMPLIFIED SCHEMATIC DIAGRAM OF A SINGLE LIMITING AMPLIFIER STAGE

Imperfections of tube construction cause slight decentering of the electron beam when the deflection voltage is zero. The circuit of Figure 4 corrects for this decentering. Deflector bias is taken from resistor voltage dividers connected to the anode voltage dropping resistors. An off-center beam produces a difference in anode currents, causing a deflector potential difference which reduces the amount of uncentering. This feedback network is bypassed to the i-f signals.

A typical four-stage limiting amplifier built using the basic circuit of Figure 4 has a limiting range of 80 db and a bandwidth of 2 mcps centered at 30 mcps. Figure 5 is a plot of output amplitude as a function of input amplitude for this amplifier. Phase stability of this amplifier was checked by the Lissajou's figure method. The Lissajou's figure was displayed on a high frequency oscilloscope having identical vertical and horizontal amplifiers. No phase change was discernible as the input signal was varied over the entire 80 db limiting range. Since phase changes of 3 to 4 degrees can barely be discerned when viewing a Lissajou's figure, the phase variations of the limiting amplifier were less than $\pm$ 3-4 degrees.

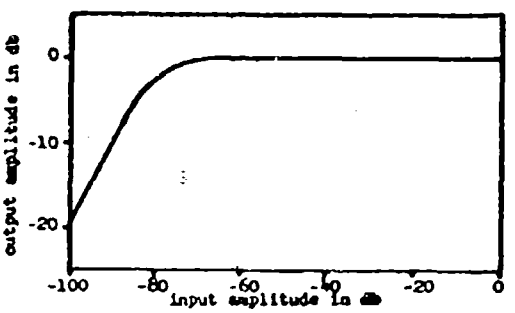


FIGURE 5 LIMITING CHARACTERISTIC OF FOUR-STAGE LIMITING AMPLIFIER

Conclusion

The beam deflection tube makes possible the construction of i-f limiting amplifiers having output amplitude and phase shift that are independent of input signal amplitude variations. These results are not feasible with other limiting devices.

A commercially available beam deflection tube, the Z2000, is about to be released by the General Electric Receiving Tube Department. Phase stable limiting amplifiers using these tubes can provide improved performance in many radar and communication systems. Examples are phase comparison monopulse radar, i-f cancellation MTI systems, several types of ECCM systems, phase modulation data transmission systems, and other applications where circuit tuning and phase shift are critical.

NOTICE: Research leading to the results presented in this paper has been sponsored in part by the Air Force Cambridge Research Center, Contract No. AF19(604)1915.

APPLICATIONS OF TRAVELING WAVE TUBES TO MICROWAVE CIRCUITS

By: Gregory E. Austin
Sylvania Electric Products Inc., Electronic Defense Laboratories

The traveling wave tube (TWT), invented nearly 20 years ago, has not yet realized its full potential as a circuit element. Many system designers are unfamiliar with the TWT's special characteristics, some of which have been discovered only recently. Because manufacturers can now produce large quantities of reliable tubes having wide varieties of specific characteristics, it is essential the the designer be aware of these characteristics so that he may apply t 'm to both wide- and narrow-band circuit . This paper reviews the use of a TWT in several applications: as an amplifier, oscillator, attenuator, circuit isolator, limiter, phase shifter, mixer, and frequency divider and multiplier.

The TWT as an Amplifier

The most important feature of the TWT is that it is a non-tunable microwave amplifier with a large, fairly uniform gain characteristic covering a wide frequency range. A typical traveling wave amplifier is shown in Figure 1.

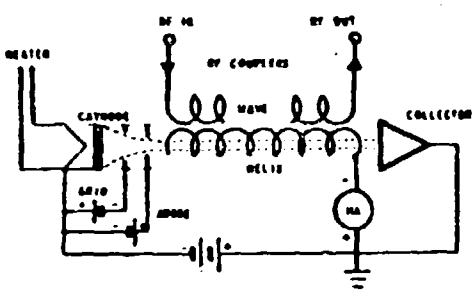


Figure 1
Typical Traveling Wave Tube Structure

The electron gun is usually comprised of an indirectly heated cathode, an accelerating anode, and a control grid. Beam density (and, in turn, tube gain) is controlled by both the positive anode-to-cathode voltage and the negative grid voltage. Although beam density is not a linear function of either voltage, tube gain is more smoothly controlled by the grid voltage. Grid voltage control curves are shown in Figure 2.

Beam velocity through the helix is determined by the cathode-to-helix voltage. Connecting the helix externally to the collector and grounding both elements increases their heat-dissipation ability, but necessitates operation of the cathode at potentials negative to the ground plane. Recently, however, low-power tubes with the helix and collector leads isolated have become available. This enables the cathode to be grounded and reduces modulator complexity on the grid.

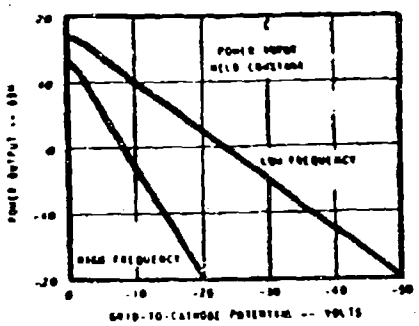
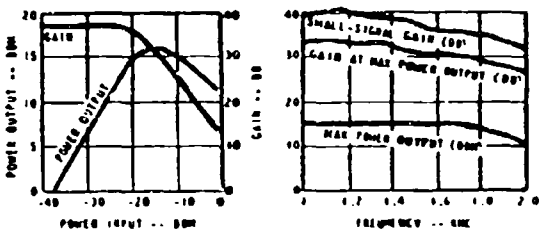


Figure 2
Grid Control

The wide-band directive couplers used to couple the rf signal on and off the helix play an important part in the TWT's broad-band operation. Their impedance is generally about 50 ohms, looking into the coupler structure. In most low- and medium-power tubes the input VSWR is less than 1.7 to 1 over the operating bandwidth; this figure can be lowered to less than 1.4 to 1 by careful manufacturing techniques.

Normally, a TWT is described by its gain, its power output, and its operating bandwidth. Figure 3 (a) shows curves of gain and power output at 1.5 kmc for a low-power tube. Small-signal gain, for instance, is characterized by the facts that the gain curve is constant and the power output of the tube is directly proportioned to the power input. Saturated gain is defined as the gain obtained when an increase in power input produces no further increase in power output; saturated gain is 6 to 7 db less than the small-signal gain. The power curve in non-linear in this section, which is known as the saturated region. At the peak of this curve, the power output is at its maximum, and for any further

increase in power input, the power relationship forms a damped oscillating curve. Never again does the power output reach the original power level. Figure 3(b) illustrates small-signal gain, gain at maximum power output, and maximum power output, over a 1-kmc frequency range for a low-power tube.



(a) (b)

Figure 3
Gain and Power Output Curves
for a Low-power Tube

It will be noted that if a strong signal is injected into the front end of the TWT, the saturation level of power output will not be exceeded. This affords good crystal burnout protection for detectors and/or crystal mixers.

Inherent Capabilities of the TWT

Because it is an amplifier, the TWT can be arranged in a circuit with a feedback path, and under controlled conditions, the result is a discrete microwave oscillator. It is a multimode oscillator, however, because of the fine ripple effect inherent in the gain structure of the tube. Regardless of this fault, the TWT is still useful as a fixed frequency oscillator.

The TWT can act as an attenuator and does act as a circuit isolator in many applications. Both of these features are inherent because of the construction of the TWT. The tube has a normal cold loss in its rf structure of about 60 db in the forward direction of propagation and 80 db in the reverse direction, due to directivity of the rf couplers. Control of the TWT from a gain of 30 db to a loss of 60 db is possible by proper bias voltage on the grid of the tube.

The TWT as a Limiter

For the first specific application of the TWT to a circuit problem, consider use of the TWT as a limiter. For example, a circuit may require a particular microwave limiter which has a broad frequency band, relatively constant power output, and a wide dynamic range of power inputs. In this case, the solution is to use two TWTs which have different power output levels.

Two tubes of the same power output level cannot be cascaded for two reasons. First, the noise of the input tube may saturate the second tube, resulting in unusable signal output. Second, cascading two tubes of the same type may limit the dynamic input range.

The limiter characteristic of the cascaded tubes depends on the saturation characteristics of both tubes. The combination of a 10-milliwatt low-power tube as the input tube and a 1-watt medium-power tube as the second or output tube, for example, will give the limiter a large dynamic range. In Figure 4 the power input versus power output curves shown for each type. They are arranged in such a way as to show the effect of a signal at one frequency operating over a range of power from -39 dbm to +5 dbm. This arrangement yields a limited characteristic of ± 2 db at a power output of 1 watt.

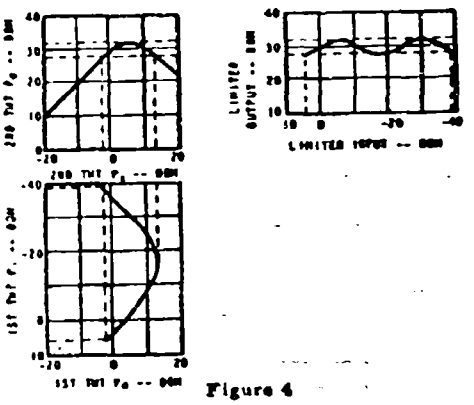


Figure 4
TWT Limiting Characteristics

A dynamic range wider than the one discussed here is possible. It was recently reported that a dynamic range of 60 db and a limited output of ±2 db was achieved using two cascaded tubes. In this case one of the TWT's was designed with characteristics which improved the saturation effect.

Special Applications

It was noted earlier that beam current and, in turn, power output, can be controlled by varying the grid voltage of the tube. The grid is useful in many applications because of its high impedance characteristic. CW signals can be converted into excellent rise-time and narrow-pulse-width signals by using the grid.

One example of this technique is illustrated in Figure 5. It shows a microwave generator which uses a klystron and two TWT's to gate an rf signal into extremely narrow pulses. In this case, the stable frequency source feeds a signal to the first TWT, which is normally biased to cutoff. A steep-

rise-time video pulse generator pulses the first TWT on. The same video signal is then delayed and applied to the second TWT in reverse polarity. The second TWT has been amplifying the signal from the first TWT, and upon application of the negative-going pulse, the second TWT cuts off the rf signal being amplified. The result of gating the two TWT's in this manner is a narrow width rf pulse having a sharp rise time. Reasonable pulse widths of this generator vary between 30 nanoseconds and 2.5 microseconds.

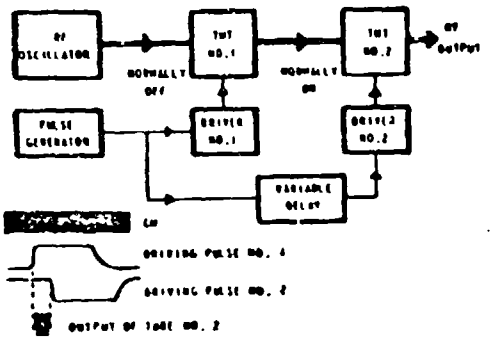


Figure 5
Nanosecond Microwave Pulse Generator

Another use of the grid is in a technique used in pulse signal control known as instantaneous automatic gain control. In Figure 6 an rf input signal is applied to a power divider circuit where half the power is fed to a TWT and the other half is detected and amplified.

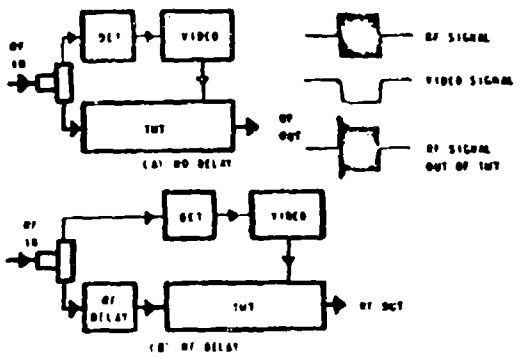


Figure 6
Instantaneous Automatic Gain Control Circuit

The output of the video amplifier, applied with the proper polarity to the grid of the tube, will control the power level of the output pulse signal. Due to circuitry delay in the detector-video amplifier-TWT grid system, the rf pulse envelope will have a sharp overshoot on its leading edge as shown in the figure. A solution would be to incorporate rf delay into the front end of the TWT (as shown in the lower part of the figure). Then, the rf pulse is delayed an amount of time equal to the video pulse delay.

Phase Shift Characteristics

An inherent problem in any microwave amplifier is the phase shift that may occur. In the traveling wave tube amplifier this phase shift may be either a problem or a means of making the tube more versatile. If accurate phase-shift tolerances are not required in an amplified rf signal, the traveling wave tube is an ideal amplifier.

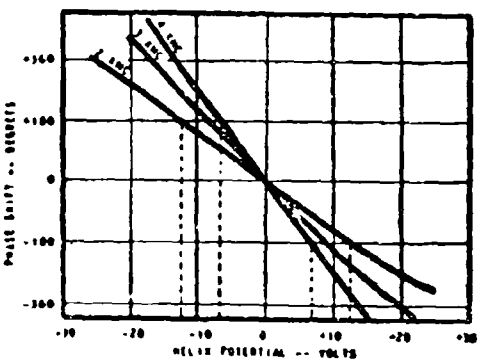


Figure 7
Phase Shift of a TWT

The curves in Figure 7 show angle of phase shift versus change in helix-to-cathode voltage. Taking the 3-kmc curve as an example, the phase in the TWT will shift 180 degrees with a 7-to 9-volt change in helix voltage. The corresponding loss in gain of the TWT does not exceed more than 5 db for this equivalent change in helix voltage. Over most of the operating frequency, the gain changes less than 3 db.

A sawtooth waveform applied to the helix of a TWT yields an approximate phase modulation of unlimited deviation. This method is called serrodyne, and has the interesting property of obtaining a single sideband signal with all other undesirable frequencies suppressed at least 20 db.

If an equally divided sawtooth voltage is applied to the helix of the tube, which is considered to have a uniform phase change, the phase of the rf signal being amplified by the tube changes from +π to 0 to -π radians, or a full 360 degrees. The sawtooth voltage is also considered to have a

short flyback time, allowing the phase to revert quickly back to its starting point. Since the period of time necessary to make a complete 360-degree phase shift is equal to the time period of the longer slope of the sawtooth waveform, the frequency of the sawtooth modulation is the translation frequency of the TWT. Therefore, for a sawtooth frequency of 1 kc, the frequency shift of the original rf will be 1 kc.

One method of studying the shift waveform and a good application of the serrodyne technique is the homodyne system. Figure 8 illustrates the makeup of a homodyne system using a TWT which is modulated by a linear sawtooth voltage. In a homodyne detection system, the carrier signal is divided into two channels. The first channel is frequency-translated, while the second is applied directly to a mixer. If the first channel is translated by a small deviation and mixed with the original carrier signal, the difference frequency recovered is the frequency of the deviation signal. In this case, the mixing process eliminates the carrier and yields only the modulated signal.

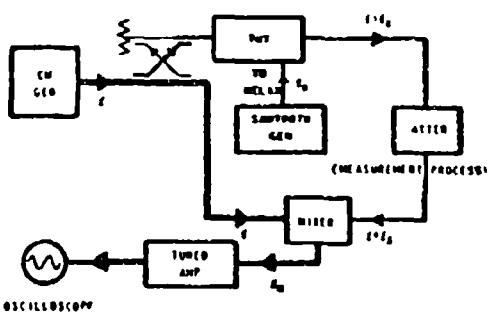


Figure 8
Homodyne Detection System

Frequency Mixing, Dividing, and Multiplying

Some recent successes in microwave frequency mixing, multiplying, and dividing have been accomplished in traveling wave tubes employing nonlinearities of an over-modulated electron beam. Conversion gains as high as 30 db and full traveling wave tube saturation power output are obtainable at microwave intermediate frequencies and other intermediate frequencies as low as 30 mc.

If a microwave signal saturates a traveling wave tube, amplification is non-linear. If another microwave signal is present, intermodulation of the two signals will occur, i.e., the TWT acts as a mixer. The IF signal can be obtained through a tuned circuit placed in the dc lead of the collector, if the frequency is low enough, or it can be obtained at the tube output if a microwave IF signal is desired.

A diagram of such a circuit is shown in Figure 9. In this circuit, the local oscillator is causing the TWT to saturate, resulting in a nonlinear bunching effect of the beam. When a signal is present on the helix, the mixing action produces a 30-mc IF current in the beam. This 30-mc signal then appears across the tuned circuit of the collector. There is one major disadvantage with this arrangement, however: there is no conversion gain.

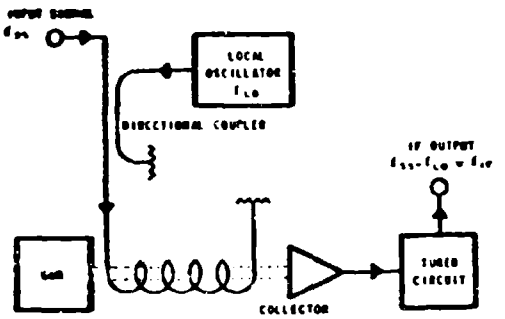


Figure 9
Frequency Mixer: IF Output from Collector

As a result of further experimentation, an improved mixer configuration was found which uses two other properties. The basis of the improved mixer is the use of a downward voltage jump and a low-voltage drift region following the amplifying helix. This is shown in Figure 10. Operation of the mixer in this manner resulted in a conversion gain of +7 db. This result is related to the gain of the amplifying section of the first helix for a low-power signal.

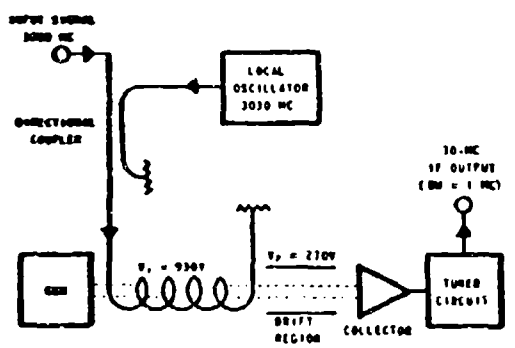


Figure 10
Frequency Mixer: Voltage Jump and Low-voltage Drift Space

A mixer with a microwave IF output can be constructed by using two helices within the same tube. The first helix is used for normal microwave amplification over a broad band of frequencies. The second helix is used as a voltage-tuned dispersive IF amplifier, operating at a microwave level. The

term dispersive means the frequency that is amplified is directly dependent on the helix synchronous voltage. The circuit is illustrated in Figure 11. For this circuit, the conversion gain was 10 db with an IF bandwidth of 20 mc. The same tube operating with a 200-mc IF bandwidth gave a conversion gain of 16 db. The saturated IF power output was +15 dbm.

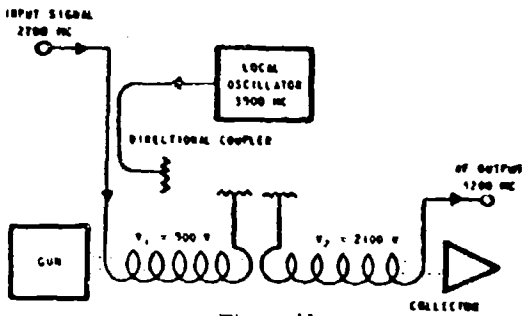


Figure 11
Double-helix Traveling Wave Mixer

An interesting application of the mixing properties of the TWT is the use of such tubes as frequency dividers employing regenerative modulation. Frequency division by a factor of two can be obtained by feedback around the circuit as illustrated in Figure 12. The input signal, f_1, mixes with a signal from the feedback path operating at $f_1/2$ and produces a mixer output $f_1/2$, which sustains the divider action. Again, the second helix is tuned to amplify the $f_1/2$ signal and suppress any signal other than $f_1/2$.

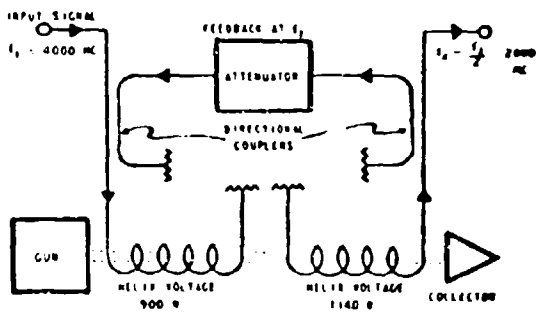


Figure 12
Regenerative Frequency Divider

A frequency multiplier circuit is the reverse of a frequency divider. A feedback path can be used, as was illustrated in the frequency divider circuit, or another method which is attractive can be used. The harmonics of the saturated signal on the first helix can be amplified by a dispersive helix following it. This will allow multiplication of frequencies without the necessity of feedback paths or local oscillators. With a wide-voltage tuning range, several different individual frequencies can be amplified by the second helix. The power level of the harmonic signal is the same as the saturated power available from the second helix.

Conclusion

When the TWT is considered for use in a microwave circuit, the operating conditions should first be defined. Each of the particular applications described in this paper will satisfy a certain function which, with care, will work smoothly. Each circuit application requires investigation to insure utmost performance, however, for a TWT is an excellent device only when used properly. The object of this paper is to acquaint the designer with some of the capabilities of this amplifier, which lends itself exceedingly well to many broad-band microwave applications. Some of the advantages of using TWTs are:

(1) For power-sensitive applications, the traveling wave tube is free from burnout caused by high level signals.

(2) The TWT, because of its broad frequency range, does not have to be tuned.

(3) TWTs can be modified easily for special application circuits.

(4) With periodic permanent magnetic focusing the TWT is attractive as a component because of its low weight and low operating power requirement.

One disadvantage, of course, is the high power required for the solenoids needed by certain groups of tubes.

Bibliography

J. R. Arnold, "Typical Operating Characteristics of Traveling Wave Tube Amplifiers," Technical Report No. 503-1, Stanford Electronics Laboratories, Stanford, California; 5 July 1957.

R. C. Cumming, "The Serrodyne Frequency Translator," Proc. Inst. Radio Engrs, v 45, p 175; February 1957.

R. W. DeGrasse and G. Wade, "Microwave Mixing and Frequency Dividing," Technical Report No. 386-1, Stanford Electronics Laboratories, Stanford, California; 12 November 1959.

"Engineering Notes, Volume I, "Huggins Laboratories, Inc., Sunnyvale, California.

F. B. Fank and G. Wade, "Traveling Wave Tube Limiters," Technical Report No. 372-1, Stanford Electronics Laboratories, Stanford, California; 31 May 1956.

"Microwave Pulse Generator, Model MPG-1," Technical Bulletin, The Hallicrafters Co., Chicago, Illinois.

R. Thornburg, "Phase Shift and Gain Characteristics of the 6493, "Sylvania Electric Products, Inc., Microwave Tube Laboratory, Mountain View, California; 8 June 1956.

RECENT ELECTRONIC SCANNING DEVELOPMENTS

by

J. Paul Shelton, Jr. and Kenneth S. Kelleher
Aero Geo Astro Corporation, Alexandria, Virginia

Summary

The purpose of this paper is to present in logical order the surprisingly small number of techniques available for achieving microsecond beam scanning in an array-type radar and to describe briefly a few of the most recent innovations in the field. The two general approaches that can be made to the problem of arbitrary coverage of a large solid-angle region are considered -- a radiating system with an output for each beam position with high-speed switching, and a single-output system with rapid phase shifting. Two general classes of multiple-beam system are discussed, one in which beam forming is done after I-F amplification and another in which multiple beams are formed by passive R-F networks. Single-input continuous scanning systems include those based upon frequency scan and those using phase shift to scan at fixed frequency. Finally, combinations of systems to provide two-dimensional coverage are discussed.

Multiple-Beam Systems

In the category of the multiple-beam systems for beam switching or constant surveillance, two lines of development have been followed. The first, for use on reception only, involves the use of separate mixers and I-F amplifiers, followed by I-F delay lines, for each element of the array, as shown in Figure 1. Multiple focused inputs are provided by appropriately tapping the delay lines in matrix fashion[1]. Since phase and amplitude control are achieved by locating the taps on the delay lines, the array configuration is not limited to linear; in fact, any convenient geometry from linear to circular may be chosen. As was previously indicated, the principle drawback to this system is its inability to operate in a transmit mode.

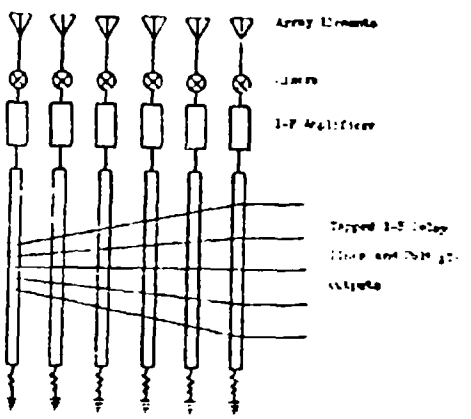


FIGURE 1. I-F PHASING MATRIX - RECEIVE ONLY

The second multiple-beam development involves passive R-F networks for providing multiple feeds to linear arrays. The Maxson system, shown in Figure 2, will shortly be incorporated in an FAA height-finding radar[2]. This multiple-beam configuration, which is of the traveling-wave type and therefore inherently slightly inefficient, provides a very convenient form factor for the waveguide application using cross-guide couplers.

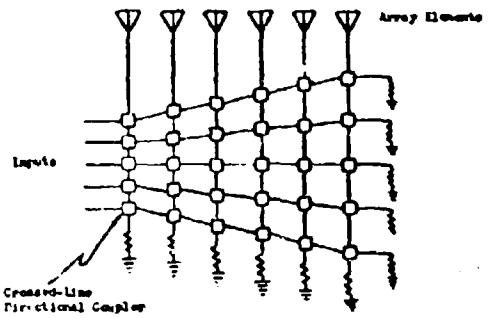


FIGURE 2. MAXSON MULTIPLE-BEAM CONFIGURATION

Although the Luneberg lens shown in Figure 3 is primarily used as a direct radiating objective with multiple-beam capability, it can be supplied

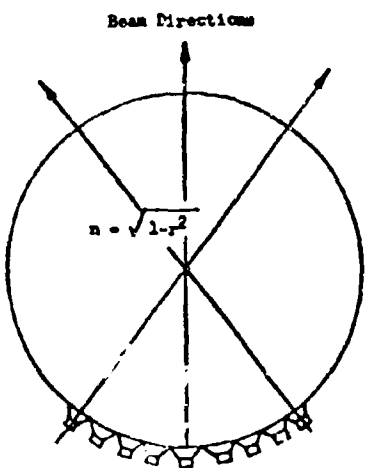


FIGURE 3. LUNEBERG LENS MULTIPLE-BEAM SYSTEM

with feeds on all sides to form a phase correction generator with which a circular or spherical array can be fed. One of the basic difficulties that is encountered in all multiple-feed objectives, such as lenses or reflectors, is obtaining sufficient beam overlap with satisfactory aperture illumination and low coupling between inputs.

The two-dimensional realization of an unusual type of linear-array feed system is shown in Figure 4. A homogeneous parallel-plate region is fed on one side by the beam inputs. The horn pickups on the other side are connected to the radiating elements by variable line lengths[3]. There are sufficient independent parameters to insure three focal points, and proper choice of feed curve introduces small wavefront distortion at other beam positions.

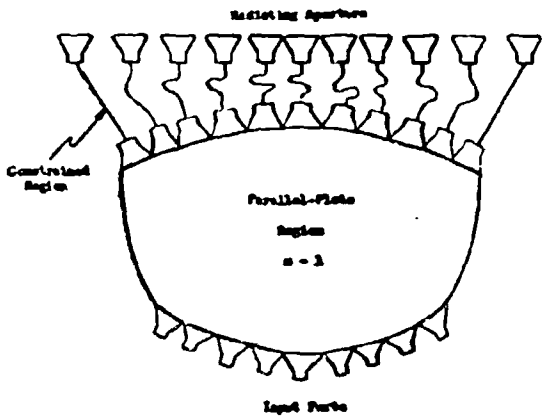


FIGURE 4. PARALLEL-PLATE CONSTRAINED LENS FOR LINEAR ARRAY

An optimum multiple-beam system is shown in Figure 5. It is built up of hybrid junctions and provides one beam position for each element of the array. This configuration is optimum in that it has isolated inputs and outputs, is theoretically lossless, and obtains maximum possible gain and minimum possible beam spacing from the array[4]. Although it does not offer the advantageous form factor of other multiple-beam configurations, it can be assembled from conventional transmission-line components.

Before leaving the subject of multiple-beam systems, it should be noted that the rapid operation of the system is dependent upon the type of switching that is used. Since many array radar systems are obtaining enormous power outputs by preceding each radiating element by an amplifier, it is possible to accomplish microsecond switching at relatively low levels, which is quite practical with existing devices.

Continuous Scanning of a Single Input Antenna System

Just as microsecond scanning with multiple feeds presumes voltage-controlled switches, fast scanning of a single-input array presumes voltage-controlled phase-shift devices. Since phase is generally dependent on frequency when a length of transmission line is involved, the advent of voltage-controlled microwave oscillators has led to frequency scan. Further, the advent of voltage-controlled solid-state devices has enabled rapid scan at fixed frequency.

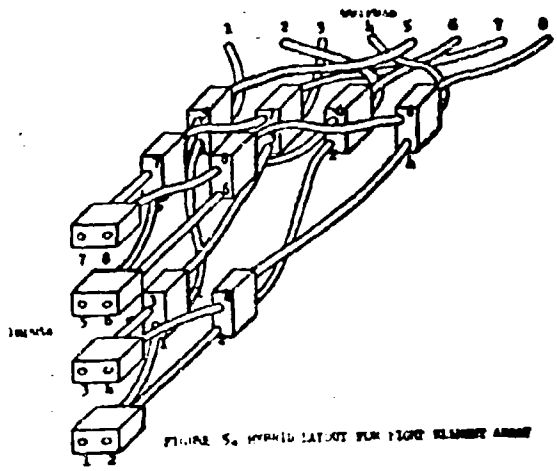
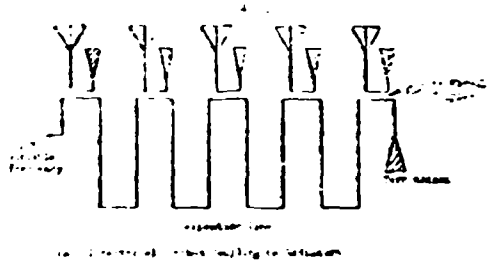
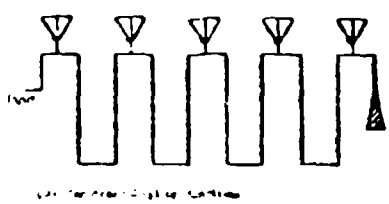


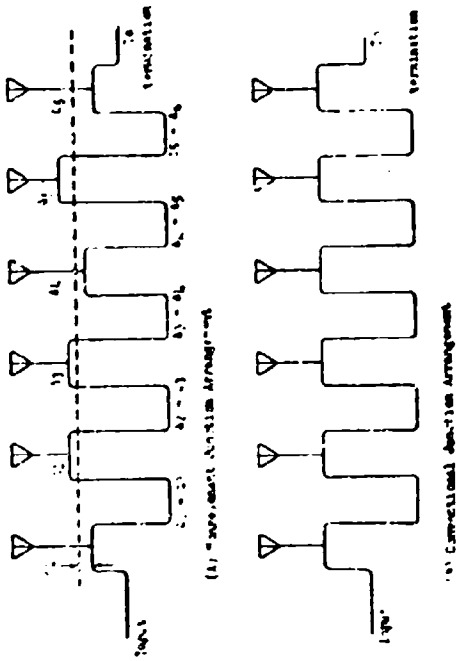
FIGURE 5. HYBRID LAYOUT FOR EIGHT ELEMENT ARRAY

Whatever the details of the frequency scan system, the basic element is a tapped R-F transmission line. Such a configuration is shown in Figure 6 for a linear array; two types of coupling junction can be used, a symmetrical three-port junction with attendant mismatch or matched





directional coupling. The primary difficulty with the simpler arrangement of Figure 6(B) is its impedance characteristic. The small reflections from the junctions can build up to a large input mismatch in a long array when the beam is scanned to the broadside position, accompanied by severe distortion of the aperture distribution. Even in the directionally coupled case the small individual VSWR's of the couplers can add to sizeable values. A solution to the problem of the broadside resonance effect is indicated in Figure 7; the overall line length from the input to the individual radiating elements is left unchanged, but the junction locations are randomly dispersed to prevent excessive mismatch at any frequency or beam position including broadside[5]. It can be shown that the input impedance for such a case has a random frequency dependence with no resonances and the amplitude distribution varies in a random fashion about the desired function. This method for eliminating broadside resonance might well be applied to antenna systems that are theoretically nonresonant, since resonances in long periodic structures can result from very small effects.



7. ELIMINATION OF BROADSIDE RESONANCE IN LINEAR FREQUENCY

The fundamental limitation of fixed array scanning antennas is the loss of effective aperture as the beam is scanned toward the plane of the array. Circular arrays avoid this problem, but the complexity of the phasing function and amplitude distribution is discouraging. A technique for obtaining 360-degree frequency scan with a circular array has recently been developed, and a sketch of the system is shown in Figure 8. Focusing is achieved by tapering the interelement line lengths in the appropriate manner to obtain a parabolic phase correction that closely approximates the required phase correction, $\frac{2\pi R}{\lambda}(1-\cos\theta)$. Amplitude distribution is controlled by using directional filters to couple energy from the serpentine line to the array[6]. Of course, any approximate focusing technique, such as the circular reflector, the Schmidt lens-reflector, the parallel-plate constrained lens system, and the frequency-scan circular array, diminishes in its effectiveness as the aperture size increases, because of the necessary reduction in usable aperture relative to overall size for given wavefront error. However, the frequency-scan circular array is found to be significantly superior to the circular reflector, and 90-degree sectors can be used to obtain beamwidths of the order of one degree, that is, with diameters of approximately 90λ.

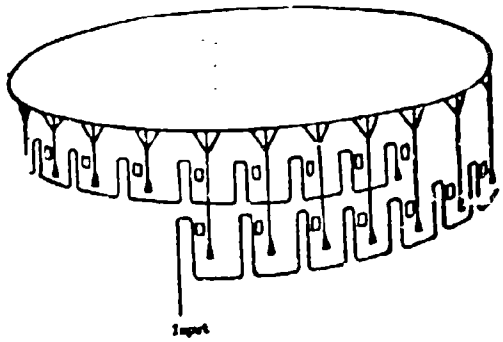


FIGURE 8. FREQUENCY SCAN OF CIRCULAR ARRAY

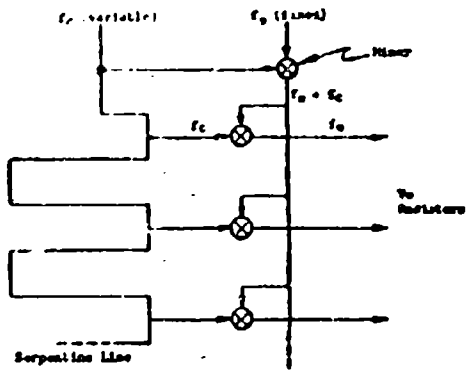


FIGURE 9. FIXED TRANSMIT-VARIABLE CONTROL FREQUENCIES

Radiation of a variable frequency is not always desirable, since the direction of the beam is a function of frequency. To retain the advantages of frequency scan with a constant radiated frequency, a network of the type shown in Figure 9 is applicable. A control frequency is passed through the serpentine transmission line to obtain the desired phase characteristic, and it is added and subtracted from the radiated frequency by mixers before and after the serpentine network.

The more conventional scanning technique for constant-frequency operation uses ferrite phase shifters or some similar voltage-controlled component as shown in Figure 10. The traveling wave configuration is generally preferred because all phase shifters can be maintained in the same state for any given beam position. The same impedance and amplitude distribution problems exist as for the linear frequency-scan array.

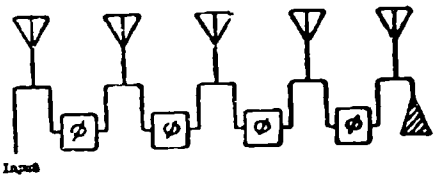


FIGURE 10. VOLTAGE-CONTROLLED PHASE SHIFTERS

The principle areas of development for voltage-controlled phase-shift devices are solid-state, which includes ferrites and semiconductors, gas plasma, and electron-beam tubes. Ferrite applications predominate at the present time, but work is under way in the other areas.

Combinations of Systems for Two-Dimensional Arrays

All of the examples in the previous paragraphs were for one-dimensional scanning, but it is apparent that the two-dimensional array that is usually used for pencil-beam scanning can combine different techniques for the two different directions of beam motion.

Almost any two of the various antenna array feed systems can be used for this purpose. It is obviously impossible to use straight frequency scan in two dimensions, but frequency scan in one dimension can be combined with any other technique. A system radiating fixed frequency is generally referred to as "phase-phase", and one using frequency scan is called "phase-frequency". It should be noted that the phase portion can be achieved by a control frequency device.

The cylindrical array represents a difficult problem for two-dimensional scanning because the phase correction of the circular array must be varied as a function of the beam position relative to the axis of the cylinder. An approximate compensation method is indicated in Figure 11. A two-dimensional Luneberg lens is used in a control-frequency phase shift mode based on the path-length focusing characteristic of the lens[7].

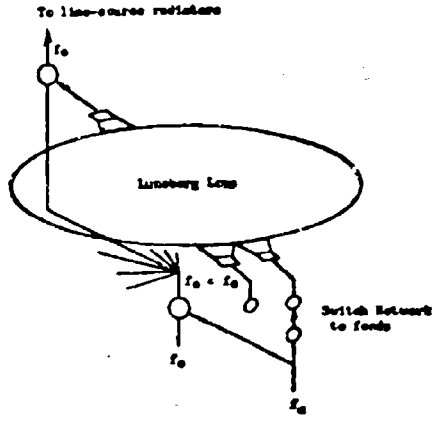


FIGURE 11. RADIUS CORRECTION IN CYLINDRICAL ARRAY

Conclusions

The techniques available for accomplishing rapid beam scanning and stepping with antenna arrays have been outlined and discussed briefly, and some recent developments have been presented.

Since a scanning time in the microsecond range is postulated, voltage-controlled operation is required. The three basic operations are switching, frequency variation, and phase shift.

Several multiple-beam configurations have been described, and the basic single-beam systems involve frequency scan and electronic phase shifters. Combinations of these are possible for obtaining volume coverage with a pencil beam.

Acknowledgment

The authors gratefully acknowledge the contributions and suggestions of Mr. A. A. Varela, who originated several of the innovations described in this paper.

References

1. To the best of the authors' knowledge, this technique was originally developed by Sanders Associates, Nashua, New Hampshire.

2. "Multidirectional Beam Scanning Antenna Array," W. L. Maxson Corp., Contract AF-30(603)1920.

3. H. Gent, "The Bootlace Aerial," RRE Journal,

Oct. 1957, pp. 47-57.

4. J. P. Shelton and K.S. Kelleher, "Multiple Beams from Linear Arrays," presented at URSI-IRE Spring Meeting, Washington, D. C.; 3 May 1960.

5. A. A. Varela and J. P. Shelton, unpublished notes on elimination of broadside resonance.

6. J. P. Shelton, unpublished notes on application of frequency scan to circular arrays.

7. A. A. Varela, unpublished notes on cylindrical array systems.

THE PULSED LIGHT THEODOLITE

By: Mr. Lee A. Jay, U. S. Army Electronic Proving Ground, Fort Huachuca, Arizona

1. INTRODUCTION

A new theodolite has joined the family of weather observation equipment used by the Meteorology Department of the United States Army Electronic Proving Ground (USAEPG), Fort Huachuca, Arizona. It is the Pulsed Light Theodolite (PLT) which may replace all present theodolites for accurate, low level wind observations because of its speed of operation and reduced man-power requirements.

For many years double theodolite observations of a balloon in free flight have been considered the most accurate and reliable of all methods of determining low level winds. Double theodolite observations are used to check other methods of wind measurement and to provide the high degree of accuracy needed for certain applications, such as determination of wind corrections for rocket firings. The results of the double theodolite method are good, but the computation work requires many man-hours of effort. Because of the rapid and frequent variations of low level winds with time, it is important to cut the time lapse between the observation and delivery of data to a minimum if the wind information is to have any value. Speed is also essential to many modern military operations.

A new system of low level wind observations is needed. The old saying, "Necessity is the mother of invention", is certainly true in this case, because the need for a new system of measuring low level winds prompted the author to consider the possibility of combining a new pulsed light range-measuring device with a standard theodolite. It seemed feasible to use the pulsed light to measure automatically the slant range to a pilot balloon while tracking it with the theodolite. This idea resulted in a contract in June 1957 with Motorola Research Incorporated of Riverside, California, for the fabrication of a pulsed light theodolite.

2. THEODOLITE THEORY AND OPERATION

The theory of operation is relatively simple. A very short pulse of light is produced from a spark in the "source" and is focused on the target. The light reflected back from the target is picked up by a collecting mirror and focused on a photo cell. The time required for the light to make the trip to the target and back determines the range (see Fig 1).

The light pulse is produced by the discharge of a 0.3 microfarad capacitor across a spark gap of the light source in the optical head. The spark exists for approximately 0.3 microseconds. The air between the spark gap electrodes is said to be heated to 100,000 degrees C. A very short and very bright flash of light results. The light is swaped into a narrow beam by two transmitter mirrors. The light reflected from the balloon is collected in the optical head by a parabolic mirror which focuses the light on a multiplier-phototube (see Fig 2). Two consecutive rangings are necessary for the computation of the wind data by the automatic computer. The values of the first ranging are stored and compared with the values obtained on the second ranging (see Fig 3). The PLT automatically produces a ranging pulse every 15 seconds. A printer accepts the height and Vx and Vy components from the computer and prints values of height in feet, the wind direction in degrees, and the wind speed in knots; e.g. 1160, 12h, 12. The number of figures in each group permits identification without confusion.

The PLT, delivered by Motorola in October 1958, consisted of the items shown in Fig 4. An inverter which could be powered by an Army Jeep battery was included but was exchanged in a short time for a small portable gasoline driven 400 cycle generator. This new generator permitted the same degree of portability and field use without the necessity of running the jeep or truck engine while operating the theodolite. This first pulsed light theodolite is a prototype or "breadboard" model to test the feasibility of the theory.

3. TEST AND EVALUATION

Tests conducted by Motorola indicated that corner reflectors would work on the balloons for reflecting light but were too expensive for regular use. A glass bead paint was selected finally, which is similar to the type used on highway signs. This is sprayed on partly inflated balloons which are permitted to dry before being deflated and saved until needed. The entire outside surface of the balloon becomes an excellent reflector with a very low percentage of scattering of the light beam.

Tests conducted by the Meteorology Department on the prototype PLT included checks of range measuring accuracy over known courses. For example, a variation of only 15 feet was found on a 2000-foot baseline. Fourteen comparative runs with a double-theodolite system were conducted. When the variations were plotted it was found that the average variation in wind speed was $\pm$ 1.64 knots, height varied $\pm$ 22.55 feet and wind direction varied $\pm$ 21.4 degrees. Most of the variation is probably due to human errors in using the two systems of manually operated theodolites. Eight tests were also conducted using a very accurate system of four photo theodolites for comparison of data. The average variation of these runs was $\pm$ 10.5 degrees in wind direction and $\pm$ 1.5 knots in wind speed. Owing to difficulties in synchronizing the timers of the two systems, height was assumed correct and was used as a base for plotting wind speed and direction.

It is believed that these comparisons indicate that the PLT can obtain winds aloft data with an accuracy at least equal to that obtained by double-theodolite methods without the loss of any computation time, since a printed tape is produced automatically as the run progresses. The important finding of the PLT test was that the idea proved feasible. The PLT was usable to slant ranges averaging 5000 feet. Maximum ranges of 7000 feet were obtained on some flights. These ranges are more than adequate for many critical Army needs. Since the PLT measures slant range it is not necessary to weight the balloons or to assume a fixed ascension rate as with normal single-theodolite observations.

4. POSSIBLE IMPROVEMENTS

Improvements possible in the next or service model will enhance the usability of the theodolite as well as its accuracy. These include the following:

a. Increase of the frequency of readout to permit the use of faster rising balloons, to provide more information on small variations in wind, and to avoid the loss of data when the balloon is missed occasionally by the light pulse, as happens with the 15-second readout.

b. Improved resolution and therefore accuracy to ± 2 degrees in wind direction.

c. More rugged and easily transportable components.

d. Provisions for self-tracking features through use of certain infrared techniques or the use of tracking helps such as variable rate motor drives.

e. Improved method of obtaining balloon reflectivity.

5. CONCLUSION

Although the FLT may pass through several development and test stages before it is a finished product, these feasibility tests indicate that it will take its place in the ever growing list of new electronic meteorological equipment.

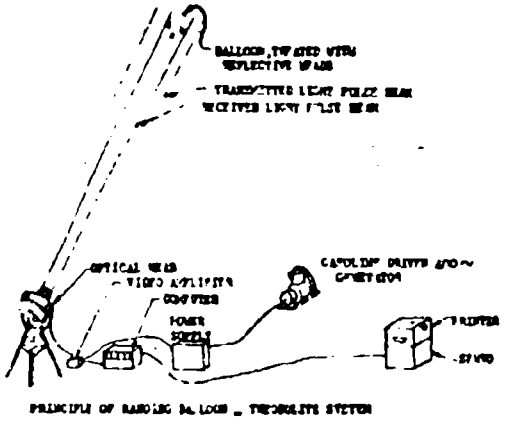


Figure 1

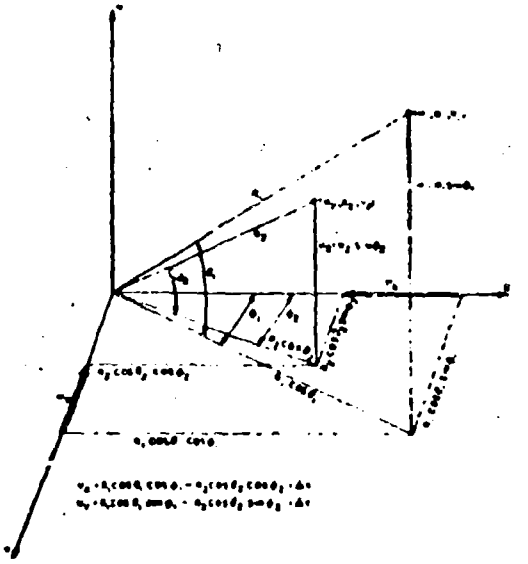


Figure 3

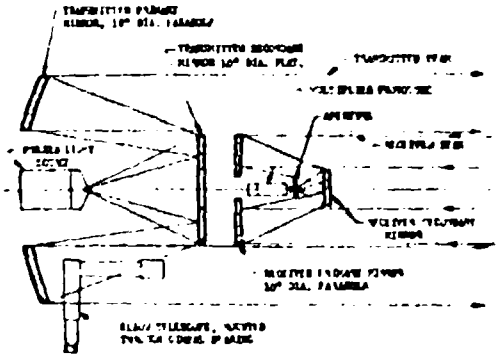


Figure 2

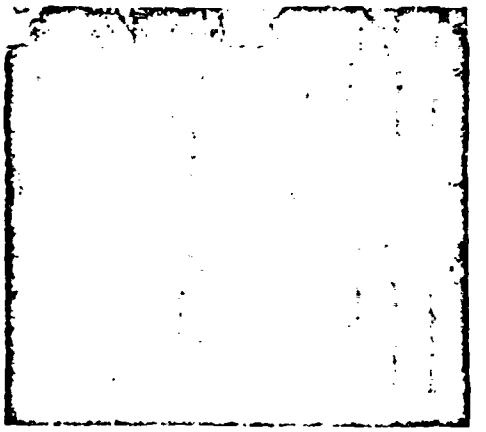


Figure 4

MULTIPLE HIGH-PRF RANGING

By: Mr. W. A. Skillman and Mr. D. W. Mooney
Air Arm Division, Westinghouse Electric Corporation

Introduction

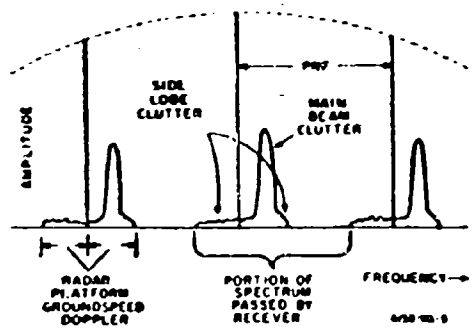
Measuring range with a low-PRF pulse radar can be done in a straightforward manner by measuring the round trip transit time of the target echo. Because a low PRF is used, the interpulse period is greater than the transit time and the echo returns before the next pulse is transmitted. In an airborne pulse doppler system, however, the PRF must generally be considerably higher, hence the interpulse period will be shorter than the round trip time of the target echo. This causes ambiguity of range measurement and allows the target to be obscured or eclipsed by the transmitted pulse at a number of ranges.

A solution to the problem is the use of multiple fixed PRFs (usually two or three) which are used sequentially. By observing the target interpulse position in each of the PRFs, the true target range can be deduced. The accuracy of such an arrangement is comparable to that of a pulse radar with an equal pulse width.

I. The Ranging Problem

The doppler effect causes a shift in the frequency of radar echos proportional to the relative radial velocity of radar and reflecting object. As a result, return from the ground to an airborne pulse doppler radar has the spectrum shown in Figure 1.

FIG.1 FREQUENCY SPECTRUM OF RECEIVED SIGNAL



The transmitted spectrum consists of discrete spectral lines centered at the transmitted frequency and spaced at multiples of the PRF. Each of these PRF lines results in a clutter subspectra centered about it as shown in the figure. Since different portions of the ground have different relative velocities to the radar, the clutter is not discrete in frequency, but is spread out by an amount of plus and minus the radar platform ground speed. Since the main advantage of a doppler radar is separation of targets and clutter on a frequency basis, a non-clutter region must be provided between adjacent clutter sub-spectra, of width equal to the maximum target speed of interest.

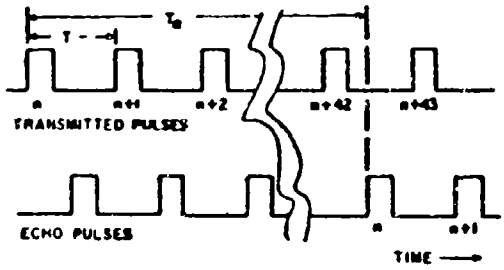
The requirement then exists that:

$$PRF \geq 2 f_{d_{radar}} + f_{d_{target}}$$

For example, at X-band, for an 1800 knot interceptor and a 1400 knot target, the minimum allowable PRF is 172 KC, corresponding to an interpulse spacing of under 1/2 mile.

The result is, of course, that a range ambiguity results for any target at greater than 1/2 mile range, as shown in Figure 2.

FIG.2 RANGE AMBIGUITIES



$T = \frac{1}{PRF}$, T_R = TWO-WAY TRANSIT TIME OF ECHO

THUS WHEN PRF IS LARGE $T_R >> T$

Since there is no way of associating the corresponding transmitted and received pulses, with a single PRF, some sort of special coding must be applied to the transmitter to permit determination of unambiguous range.

The most important factors in selecting a method of coding the transmitter are the avoidance of spreading the clutter spectrum, and generating spurious signals. The coding for example, could take the form of carrier FM, continuously varying PRF, or pulse jitter. These all, however, cause spreading of the clutter due to the modulation. Also in cases where the PRF must be variable, spurious signals often arise from high order harmonics of the PRF.

Obviously, it is highly desirable to use stable crystal-controlled oscillators for PRF generation to minimize these effects. By selecting one of several fixed values of PRF sequentially, and deadening the receiver until switching transients die out, a system is obtained with the same clutter rejection and spurious free performance as a single PRF system.

There are several other advantages to such an arrangement, the most important of which, to a single target tracking radar, is that a method of tracking through an "eclipse" is provided. Eclipsing is the obscuration of the target by subsequent transmitted pulses, which occurs when the target has a range corresponding to multiples

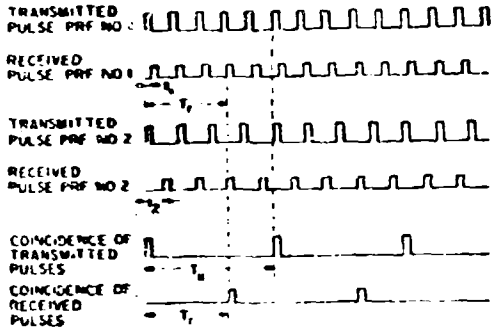
of the interpulse interval. With the mechanization to be described, an eclipse can be anticipated, and before it occurs, the PRF can be switched to one of the other available values which is certain to prevent an eclipse. Therefore continuous eclipse-free tracking is obtained.

The accuracy of a multiple PRF system is comparable to a single PRF radar with the same pulse width, since the multiple PRFs merely resolve the basic ambiguity, while tracking is done on the same pulse basis as a pulse radar.

II. Two PRF Ranging

The principles involved in a two-PRF ranging system are shown in Figure 3.

FIG.3 PRINCIPLES OF TWO-PRF RANGING

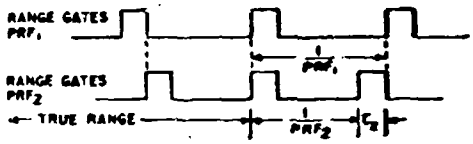


In this system, the ambiguous range to the target is determined first in PRF No. 1, and then the ambiguous range is re-determined in PRF. No. 2. By taking a time comparison of the two ambiguous ranges, made possible by synchronism of the PRFs, a coincidence pulse is obtained which is a measure of the true range.

As shown in the figure, the PRFs are chosen to have a common sub-multiple frequency. If the transmitted pulse trains are then compared in a coincidence detector, the common submultiple frequency is obtained. Similarly, if a range track is established on the target first at one PRF and then the other, comparison of the range gates gives the same submultiple frequency, but shifted in time. By measuring the time delay between the two sets of coincidence pulses, the true target range is obtained.

The relations between the various parameters which are required to obtain this desired operation, will next be considered. The common submultiple frequency is chosen so that its interpulse period is equal to, or greater than, the expected target range. The minimum value of PRF is chosen as described above. The ratio of the PRF to the ranging frequency is set equal to an integer, N, by adjustment of either frequency. The second PRF is then set equal to (N + 1) times the ranging frequency. Other values for the second PRF are possible but this value gives the widest allowed pulse width and closest PRF values. The transmitted pulse width has a maximum allowable value, τ_2, depending on the PRF values and the required resolvable range. Figure 4 is an enlarged view of the range gate pulses in the immediate vicinity of the target true-range coincidence

FIG.4 MAXIMUM ALLOWABLE PULSE WIDTH



① $\frac{1}{PRF_1} = \frac{1}{PRF_2} + \tau_2$

② SUBMULTIPLE FREQUENCY =
$f_0 = \frac{PRF_1}{N} = \frac{PRF_2}{N+1}$
WHEN N IS AN INTEGER

③ SOLVING 1 AND 2
$\tau_2 \leq \frac{1}{N(N+1)f_0} = \frac{1}{(N+1)PRF_1}$

Clearly, to prevent a false coincidence at other ranges than the true one, the pulses adjacent to the true coincidence must not overlap, as shown in the figure. The resulting maximum pulse width is seen to be dependent on the range to be resolved, and on the ratios of the PRFs to the ranging frequency.

From the standpoint of obtaining the maximum range performance from the radar, a high average transmitted power is desired. This implies a pulse width as wide as possible to prevent the requirement for excessive peak power. Consider, as an example, the system which requires PRFs in excess of 172 KC, and assume that an unambiguous range of 50 miles is required. This corresponds to an f_0 of 1.62 KC and an N of $\frac{172}{1.62} = 106$. The maximum pulse width is:

$$\tau_2 = \frac{1}{106 \times 107 \times 1620} \approx 0.05 \text{ microseconds}$$

This would require a rather high peak power to obtain any reasonable average power, and is especially undesirable when the high PRF is considered, since the modulator problem becomes severe. The solution to this dilemma is the use of three or more PRFs as will be discussed next.

III. Three PRF Ranging

Basically, the use of three PRFs for ranging may be considered as the two PRF method used twice. The three PRFs may be thought of as forming two pairs of PRFs, each pair measuring range to a different sub-ranging frequency. These two sub-ranging frequencies then form, in turn, a two PRF ranging system which measure range to the final ranging frequency.

Many different possibilities arise in the choice of the integers used in the three relations. One choice which leads to a relatively simple mechanization will be discussed. As shown in Figure 5, the equation for the ranging frequency, f_0, is

$$f_o = \frac{PRF_1}{N(N+1)} = \frac{PRF_2}{N(N+2)} = \frac{PRF_3}{(N+1)(N+2)}$$

FIG.5 THREE PRF RANGING EQUATIONS

① SUBMULTIPLE FREQUENCY =

$$f_o = \frac{PRF_1}{N(N+1)} = \frac{PRF_2}{N(N+2)} = \frac{PRF_3}{(N+1)(N+2)}$$

② MAXIMUM PULSE WIDTH =

$$\tau_3 = \frac{1}{N(N+1)(N+2)f_o} = \frac{1}{(N+2)PRF_1}$$

③ LOWEST COMMON MULTIPLE FREQUENCY =

$$f_r = N(N+1)(N+2)f_o$$

It is seen that the lowest common multiple, f_r, of all three PRFs is

$$f_r = N(N+1)(N+2)\,f_o$$

The PRFs are approximately N^2 times the ranging frequency, so the value of N is roughly the square root of that necessary in a two PRF system. Comparison of the equations for pulse width show that this is increased by the square root of the value of N for two PRFs also.

In a manner similar to that used for the two PRF system, the maximum allowable pulse width, τ_3, can be shown to be

$$\tau_3 = \frac{1}{N(N+1)(N+2)\,f_o} = \frac{1}{(N+2)\,PRF_1}$$

With three PRFs and the values of PRF (172 KC) and ranging frequency (1.62 KC) used in the previous example, N now becomes 10.

$$\tau_3 = \frac{1}{10 \times 11 \times 12 \times 1620} = .5\ \mu sec$$

Thus, ten times the average power may be obtained using the same peak transmitted power, if three PRFs are used instead of two. Generation of the wider pulse by a modulator is considerably easier. These advantages are partially offset by the fact that three detections of the target are required, which has lower probability of success than the two detections required of a two PRF system, and also more time is required.

IV. Mechanization

A block diagram of one possible mechanization of a three PRF range measuring system is shown in Figure 6. This makes use of the relations discussed previously. Since all three PRFs have a common multiple, f_r, dividers are used to generate each PRF from a crystal oscillator at f_r. Thus, all PRFs are automatically phase locked. Any one of the PRFs can be used to key the transmitter by use of the PRF selector.

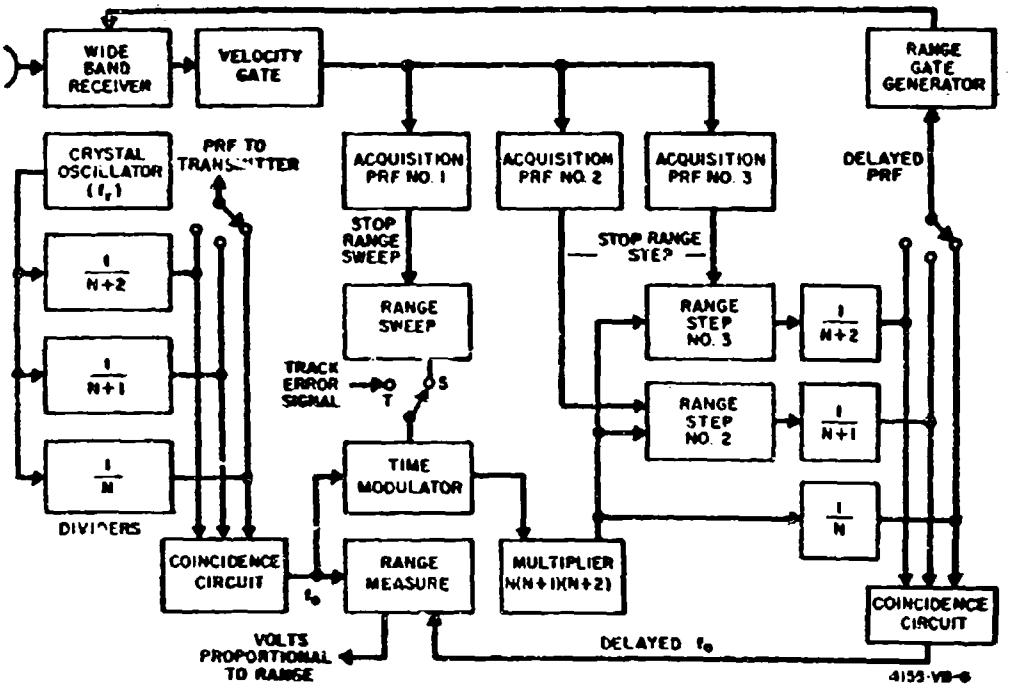
All three are fed continuously to a coincidence circuit which provides pulses at the ranging frequency, f_o. This will be used as the time base for measuring range, much as the transmitter PRF is used in a pulse radar. In order to generate range gates which can move through many ambiguous range intervals to track a target, f_o is put into a time modulator which delays the pulses by a variable time, controlled either by the range sweep circuit or the tracking circuit. It should be noted that the delayed pulses from the time modulator will not, in general, represent the true range to the target. These pulses are next multiplied in frequency by $N(N+1)(N+2)$ to obtain f_r. Pulses at this frequency are then divided down to the PRFs in the same manner that the transmitted PRFs are generated. Again, the outputs are put into a coincidence circuit which provides pulses at the ranging frequency, which, after range measurement, are delayed by the true range from the coincidence of the transmitter pulses. The two sets of pulses are used to generate a voltage proportional to range in the Range Measuring circuit. The outputs of the dividers go to another PRF selector which is synchronized to the first, to apply range gates to the receiver.

The actual measurement of range occurs as follows. PRF_1 is transmitted and the range gates are searched by means of the range sweep and time modulator. Upon detection of the target, the range sweep is stopped and the time modulator maintains its delay. PRF_2 is now transmitted and the range gates again searched. This search is done by the range step circuit. By omitting one pulse to PRF_2 divider the range gates are caused to jump an interval which is the reciprocal of f_r. Thus, the interpulse interval may be searched by continuing to omit pulses at the desired step rate. Upon detection of the target, this procedure is stopped, thus leaving the counter at the proper phase to position PRF_2 range gates over the target. The same procedure is used in PRF_3 to obtain the third ambiguous range measurement. At this time, since all range gates line up with the target, the true range is measured and is available as a voltage as previously discussed.

By moving the range gate back and forth across the target a small amount, called jittering, an error signal may be developed. This is used to move the time modulator output pulses in such a way as to center the gates on the target to provide continuous tracking.

An eclipse detector circuit compares the transmitter pulses with the range gates to determine when they are about to become coincident, or eclipse. The PRF is then switched to one of the other two values available, thus maintaining continuous tracking.

FIG.6 THREE PRF RANGE MEASURING CIRCUITS



RECENT ACHIEVEMENTS IN MISSILEBORNE MAGNETIC RECORDERS

By: Mr. Mark M. Siera, Lockheed Missiles and Space Division

REQUIREMENTS FOR MISSILEBORNE MAGNETIC RECORDERS

The most important differences between magnetic recorders used in other industries and those for use in airborne vehicles are in weight, size, and power requirements. For airborne vehicles, all these factors should be as low as possible – but the data storage capability must be as great as possible. These requirements are not very compatible. Furthermore, as a generalization, environmental specifications call for operation of missileborne magnetic recorders in an ambient temperature range of -30° to +200°F; exposure to shocks up to 100 g; high acceleration; and vibration at frequencies of 5 cycles to 2 kc with peak deviations equivalent to 20 g or more at 100-percent humidity, and of course, from normal pressure of 760 mm to 10^{-9}mm or almost absolute vacuum. In addition, the recorders should have minimum flutter, no amplitude variations, high linearity, and a signal-to-noise ratio of 35 db or better. (The word noise in the term signal-to-noise ratio refers to anything that is not data; it includes, then, flutter and wow, amplitude deviations, dropouts, etc. For example, a signal-to-noise ratio of 40 db means that all the inaccuracies combined should never exceed 1 percent, 34 db, 2 percent, etc.)

RECENT DEVELOPMENTS IN MISSILEBORNE RECORDERS

The Lockheed missileborne recorder AMR-100, which was developed in 1958, is shown in Fig. 1. This

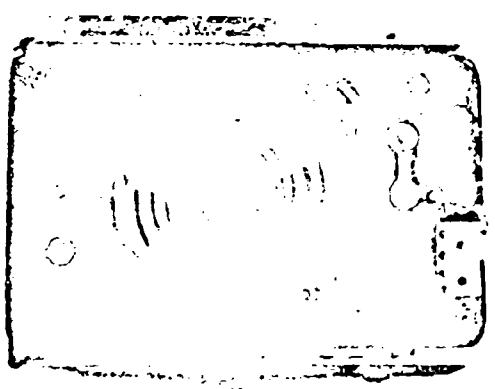


Fig 1 Lockheed Missile System Division's AMR-100 (Courtesy of LMSD)

unit weighs 6 lb, has a volume of 235 cu in., and can store 3 million bits of information on two recorded tracks. Instead of the conventional two-reel configuration, an endless loop tape magazine (in center of Fig. 1) is used. Up to 450 ft of tape can be contained between the two drive plates, the upper of which is invisible in the illustration. As can be seen, these plates have a helical indentation which allows the tape to slip toward the center in order to go from the speed at the outside of the endless loop to that on the inside. The two plates are driven by a pulley whenever the rubber idler is engaged to it. The tape is taken off at an angular post, guided over elevated rollers, then turned into an attitude parallel to that of the heads, and brought out past the input guide roller, past the erase and record/reproduce head. Driven by a high precision capstan, it then passes the output roller and goes back into the magazine and into the outside layer of the endless loop.

More recently the Jet Propulsion Laboratory developed the smallest video recorder the author has seen (Fig. 2). The diameter of the base plate is



Fig. 2 Jet Propulsion Laboratory's Video Recorder (Courtesy of JPL)

7-1/2 in., and the total weight is less than 1-1/2 lb. Its power consumption in the reproduce mode is in the neighborhood of 100 mw. The term video recorder usually suggests bandwidths up to several megacycles; however, the JPL recorder can record and reproduce still pictures only. Reproduction of the still pictures is done by superimposing one still picture after the other, each with very low definition because of extremely low frequency response. The bandwidth during recording is only 10 cycles to 10 kc, while the reproduction takes place at a much slower speed, having a bandwidth of 0.01 cycles to 10 cycles only. By very accurate control and high accuracy and precision of the mechanical components, the original bandwidths can be restored by reproducing lower bandwidth signals, 50 or 100 times, and then transmitting these signals and superimposing these data to receive, finally, a picture as shown in Fig. 3. This figure,

---

Excerpts of this paper were previously presented to the LMSD Seminar, 24 September 1959, and the Convention of the AES, 8 March 1960.

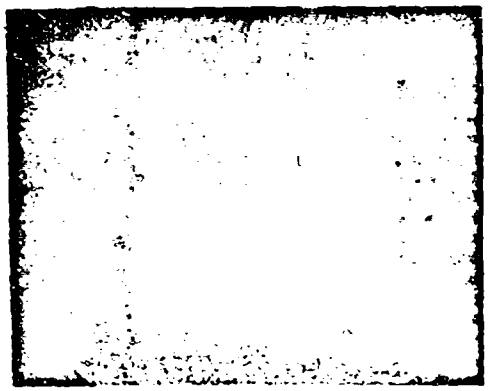


Fig. 3 Lunar Crater Reproduced by Jet Propulsion Laboratory's Video Recorder (Courtesy of JPL)

showing a lunar crater, was reconstructed after 50 reproduce cycles and 50 transmissions.

In the course of the studies by the Jet Propulsion Laboratory, many systems were evaluated, and many approaches are continuously investigated for use in missile recorders. One of these is a flux responsive system, the reproduce head of which is illustrated in Fig. 4. This head is built in a bridge configuration;

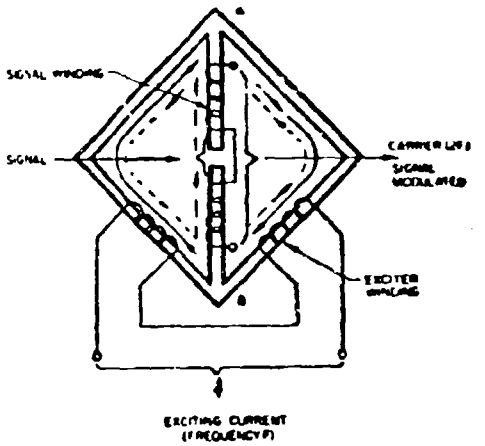


Fig. 4 Flux Responsive Reproduce Head

as the tape containing the signal passes the gap, it unbalances the equilibrium of the bridge (a process described years ago by Otto Kornei of IDM). The head is constructed in such a way that an exciting current of the frequency F is induced into the exciter winding on two arms of the bridge, while a signal winding is wound onto the other two arms of the bridge. Any magnetic field that is present at the gap will then modulate the second harmonic of the exciting current or carrier, which can then be demodulated and amplified so that the data can be used. Such a head has a particular advantage for the reproduction of information stored in missile recorders: the old Faraday formula $E = N(d\phi/dT)$ which describes the dependence of the reproduced voltage from the rate of flux change does not enter into this flux responsive method at all, which means that the speed at which the recorded tape passes the gap is of little or no importance and reproduction of recorded data is possible even while the tape is stationary over the head. This feature is, of course, very attractive in processes where the size of the unit and especially the length of the tape are limited.

More than 20 years ago a rotating magnetic head was invented in Paris. This magnetic head is still used by several manufacturers of television recorders because it permits using a maximum amount of magnetizable tape at very low tape speeds; use of a fast-rotating head results in a high rate of change in flux, which causes a high packing density or frequency response. A development of the Armour Research Foundation is depicted in Fig. 5. On a common pole

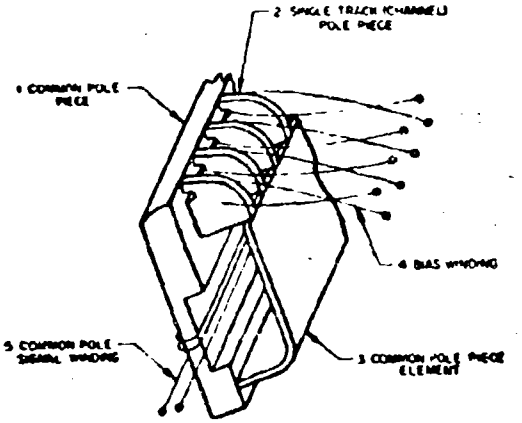


Fig. 5 Common Pole Record Head

head is a common pole piece (No. 1), onto which the common pole signal winding (No. 5) is wound. Separated from it by the active gap are many single-track pole pieces (No. 2), which in turn carry individual bias windings while being magnetically connected to the common pole piece element (No. 3). This system is designed to make use of the fact that AC bias moves the recorded data up and down on the hysteresis loop—an effect that is used in direct recording in order to increase the linearity (i.e., to remove the recorded information from the curved part of the hysteresis loop into its linear region). Depending upon the possible limitation of the minimum recorded data level and the necessary signal-to-noise ratio, one can by exaggeration reach the conclusion that there is no recording without bias and vice versa. This common pole recording head can then be used with a result similar to that achieved with the rotating head if a signal is continuously present on the common pole signal

winding while the bias is switched from one single-track pole piece to the other; therefore, during relatively slow advance of the tape across the gap, only small areas of the tape will be recorded, depending upon the physical dimension of the bias track and the fast rate of bias switching. Development of this head may lead to very interesting applications.

A magnetic recorder for the first manned satellite, Mercury, was developed by the Consolidated Electrodynamics Corporation. (See Fig. 6.) This

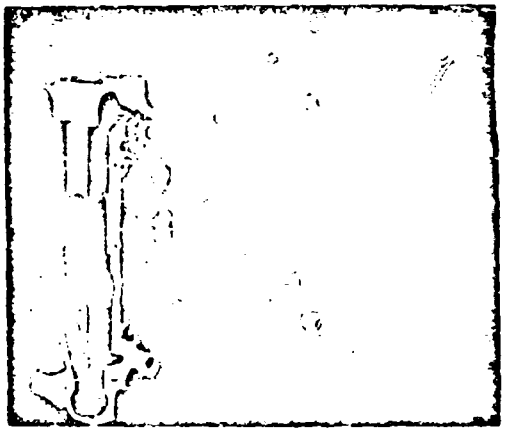


Fig. 6 Satellite Recorder 5-721 Developed by Consolidated Electrodynamics Corporation's Data Lab (Courtesy of CEC)

recorder does not reproduce; since the man must be recovered, the records will be recovered also, and the data will be reproduced after recovery. This recorder is intended to record physiological and aeromedical reactions of the man in space, as well as environmental conditions, temperature, pressure, accelerations, and shock during a maximum period of 8 hours. The recorder has a tape speed of 1-7/8 in./sec with a flutter and wow in the neighborhood of 1-percent peak to peak from dc to 500 cycles. There are 7 tracks on 1/2-in. tape, carried on 10-1/2-in. reels. The reels themselves and all the components and mounting surfaces (the light-colored parts in the figure) are of magnesium alloy, which accounts for a total weight of less than 12 lb. The total power requirement is approximately 12 w. A view of the packaging of the electronic components is shown in Fig. 7. The size of the whole recorder is 13 by 13 by 3-5/8 in. Its frequency response is in the audio range.

Figure 8 shows a development of the Cook Electric Company of Chicago. This recorder, model DR 25-2, was developed for the re-entry phase of an ICBM nose-cone flight. It is 7 in. long, has a diameter of 5.34 in., and weighs less than 7 lb. This unit also records only; the data will be reproduced when it is recovered. The recorder contains 900 ft of 1/4-in. tape and can record continuously for 4 min on two tracks at 45 in./sec. Its flutter and wow specifies less than 1 percent rms from dc to 300 cycles, and its frequency response is ±3 db from 1 kc to 60 kc.

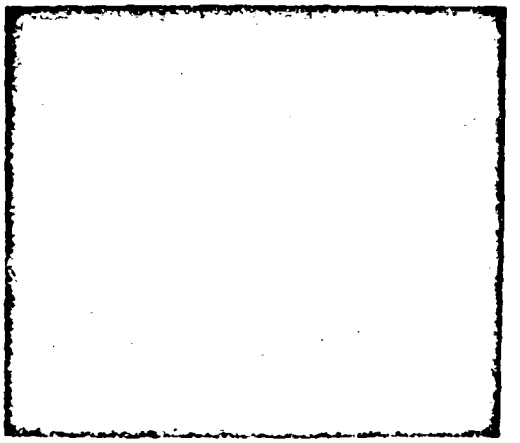


Fig. 7 CEC Data Lab's Satellite Recorder 5-271 With Electronics and Drive System Exposed (Courtesy of CEC)



Fig. 8 Cook Electric Company's ICBM Recorder DR 25-2 (Courtesy of Cook Electric Co.)

Another development by the Cook Electric Company is the miniature airborne record/reproduce system, DR 24-2, shown in Fig. 9. The rear of the recorder is pictured in order to indicate that magnetic heads can be used for various purposes; in this case they are used for speed control. The recorder measures 8 by 6 by 7 in. and weighs 11 lb. It contains 75 ft of 1/2-in.-wide tape, which is run at 40 in./sec. Power consumption is less than 85 w. The unit has four tracks with a response of ±1 db from 800 cycles to 50 kc, and a wow and flutter of 0.4 percent rms from 0 to 300 cycles.

Figure 10 shows a product of the Leach Corporation – a recorder without playback facilities – which

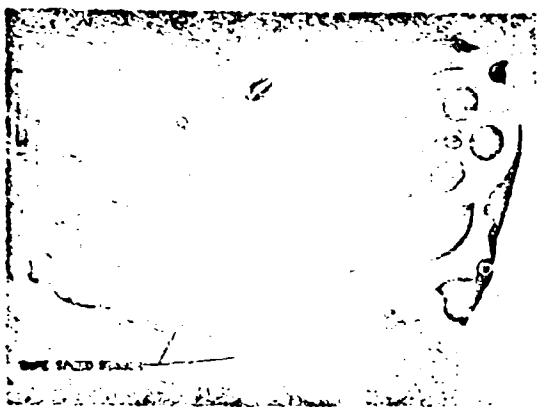


Fig. 9 Cook Electric Company's Airborne Miniature Recorder DR 24-2 (Rear View) (Courtesy of Cook Electric Co.)

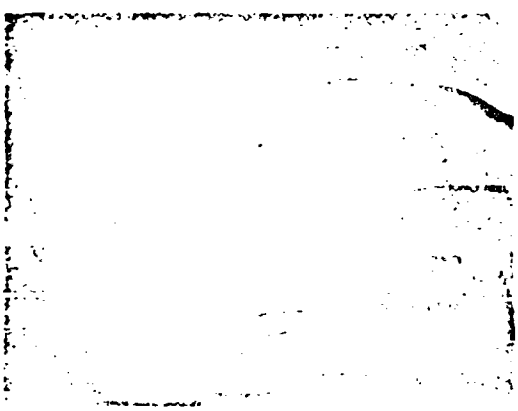


Fig. 10 Leach's Missile Recorder MTR-362 (Courtesy of Leach Corp.)

is now in production. The unit is produced with up to 14 channels of analog or digital recording capability. Most applications, however, use a carrier erase system whereby a 1000-cycle signal is prerecorded onto the tape, and the record heads are directly connected to the transducers so that data frequencies from 0 to 250 cycles erase the prerecorded frequency at the rate of the measured data.

Another interesting missileborne recorder is the Leach Model MTR-1200, shown in Fig. 11. This recorder can store up to 650 ft of tape and operates at 60 in./sec. It is used for recording standard IRIG FM/FM channels. Exact information on capabilities, physical size, and power requirements was not available.

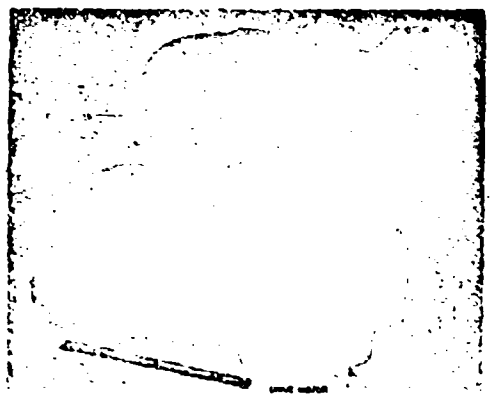


Fig. 11 Leach's Missile Recorder MTR-1200 (Courtesy of Leach Corp.)

The external configuration of the Model 003-3 recorder produced by the Speidel Corporation, Providence, R. I., is shown in Fig. 12. Its

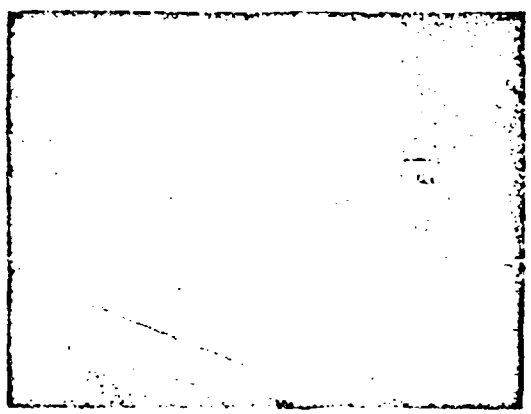


Fig. 12 Speidel's Missile Recorder 003-3 (Courtesy of Speidel Corp.)

dimensions are 2 by 8 by 6-1/2 in., and it weighs 10-1/2 lb. Figure 13 shows the arrangement of its components. The tape speed is variable, from 58 to 66 in./sec. At its normal working speed of 60 in./sec., the unit records and plays back FM/FM data in the range of 1 to 100 kc. Wow and flutter is 1 percent rms from dc to 5 kc. The total power requirement is approximately 56 w. The transport system and endless tape loop magazine used in the unit are shown in Fig. 14; this whole section is normally covered by a pressure cover which seals the unit in order to retain a liquid in which the endless tape loop operates. It is obvious that the presence of this liquid reduces the high-frequency response because of (1) separation of

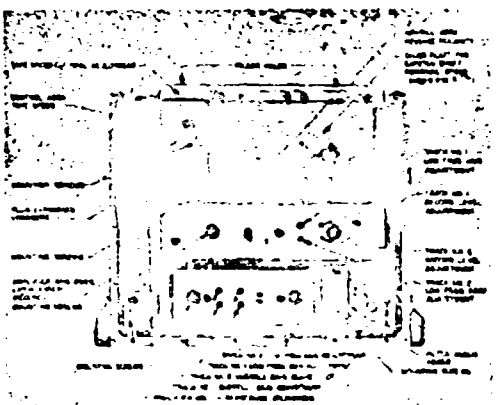


Fig. 13 Electronics of Speidel's Missile Recorder 003-3 (Courtesy of Speidel Corp.)

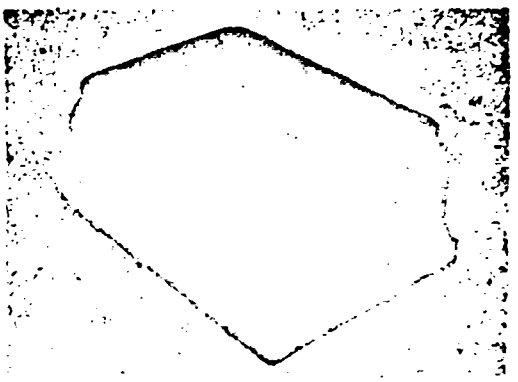


Fig. 15 LMSD's Satellite Recorder AMR-104 (Courtesy of LMSD)

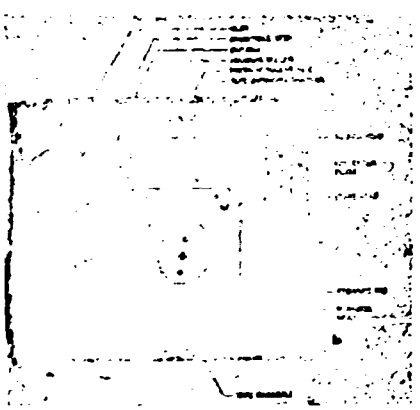


Fig. 14 Drive System and Endless Loop Magazine of Speidel's Missile Recorder 003-3 (Courtesy of Speidel Corp.)

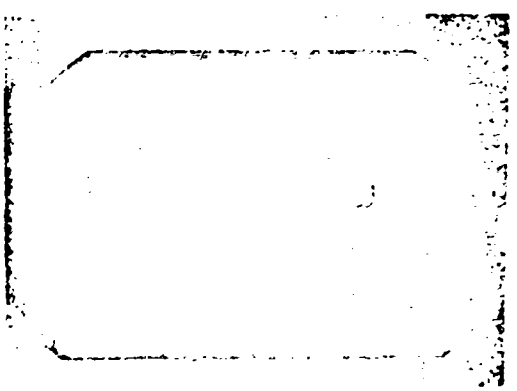


Fig. 16 LMSD's Satellite Recorder AMR-104 With Cover Removed (Courtesy of LMSD)

the tape from the record and (2) playback gap. The manufacturer has reduced this effect somewhat by notching the playback head so that lubricant particles are wiped off just before the tape reaches the playback gap. however, this increases the contour effect to such an extent that the low-frequency response is limited and the minimum reproducible frequency is 1 kc. On the other hand, remarkable gains are achieved by use of the liquid, which not only lubricates but dampens high shock effects, permitting the recorder to be used under extreme environmental conditions.

A recent development of the Lockheed Missiles and Space Division, the airborne magnetic recorder AMR-104, is shown in Fig. 15. The unit has a volume of less than 275 cu in. and weighs approximately 10 lb. It is hermetically sealed and fulfills all the rugged requirements specified previously. Figure 16 shows the unit with the top and magazine covers removed. In the foreground are the high-precision components. At the front is the erase head, the rollers of which belong to a closed-loop configuration (as in precision ground recorders). It is seen that the tape goes from the input roller to the capstan, passes the record head, goes over the turnaround idler and the reproduce head, returns to the capstan, and then goes to the output roller and magazine. The magazine contains up to 800 ft of tape in an endless loop and is very similar to the unit shown in Fig. 1. The Lockheed unit is capable of recording 3 hours of data continuously and reproducing the data within 5 min. The record-to-playback speed ratio is 1:50, the record speed is 0.8 in./sec, and the reproduce speed is 30 in./sec.

Figure 17 shows the inside of the recorder AMR-104. The motor is 3-phase, 400-cycle, single-speed, with a special two-clutch arrangement to produce the 50:1 speed ratio. There are two different worm

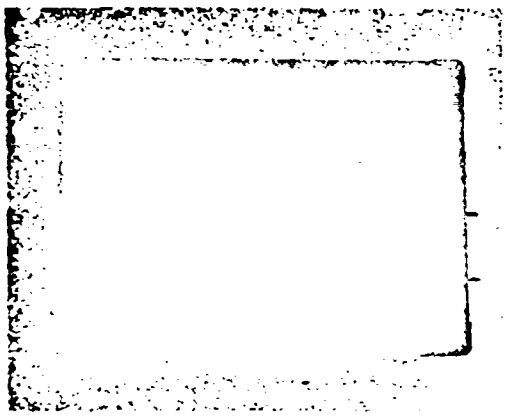


Fig. 17 Bottom View of LMSD's Satellite Recorder AMR-104 (Courtesy of LMSD)

gear configurations on the clutches; while one clutch is slipping, the other clutch is driving in the other direction, so that the high-speed ratio is produced simply by changing the phase sequence to the motor, using two relay contacts. The lower part of the figure shows the packaging of the electronic components on four printed circuit boards which include the necessary electronics for two tracks and all the relays for the control functions. The package contains, for each track, an AM modulator, a record amplifier, a bias oscillator, a playback amplifier, and an AM demodulator. The unit can record data from dc through 1250 cycles; the highest reproduce frequency at 30 in./sec is 72 kc. Total flutter and wow of the unit is less than 1/4 percent rms from dc to 1 kc, and the signal-to-noise ratio is better than 35 db.

Another Lockheed unit, the AMR-106, is still being developed. A square wave reproduced on this unit is shown in Fig. 18. The frequency of this wave

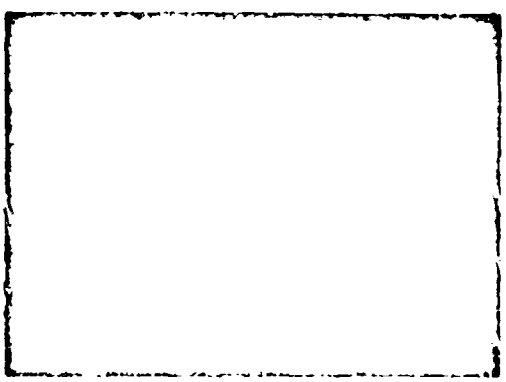


Fig. 18 250-kc Square Wave, Recorded and Reproduced on LMSD Satellite Recorder AMR-106 at a Tape Speed of 30 in./sec. (Courtesy of LMSD)

is 250 kc while the record and reproduce speed is only 30 in./sec. The method of achieving this high response will be disclosed in the near future.

FIGURE OF MERIT CALCULATION FOR MISSILE-BORNE MAGNETIC RECORDERS

This section describes a figure of merit calculation which is advantageous for missileborne magnetic recorders but could also be used for other recorders. As in any figure of merit calculation, the accuracy is not very high, and not all parameters which might have to be considered are introduced. The calculation is as follows:

C = information storage capacity in bits
B_R = record bandwidth in cps
T_R = record time in seconds
N_T = number of parallel tracks
S/N = rms signal to noise ratio in db

Shannon's formula for information capacity:

$$C = B_R \, T_R \, N_T \log_2 \left[1 + (S/N)^2\right] \text{ bits}$$

Strictly speaking, Shannon's formula applies where N has gaussian statistics; it assumes somewhat different forms for other types of noise. However, if S/N is much greater than unity – as is usually the case – all the forms can be written approximately as follows:

$$\begin{aligned} C &\approx 2 \, B_R \, T_R \, N_T \log_2 (S/N) \\ &= 6.64 \, B_R \, T_R \, N_T \log_{10} (S/N) \\ &\approx \frac{B_R \, T_R \, N_T}{3} \, (S/N) \end{aligned}$$

The following physical characteristics are now introduced:

V = volume in inch^3
W = weight in lb
P = power in watts

The figure of merit is then:

$$\frac{C}{VWP} \quad \text{or} \quad \frac{B_R \, T_R \, N_T}{3VWP} \, (S/N)$$

This is, of course, as arbitrary as the figures of merit on vacuum tubes and transistors, but it might serve as a tool for making a rough comparison of recorders, as long as the user does not forget that many criteria have to be introduced or neglected, depending upon the usage of the recorders.

Table 1 gives a comparison of the figures of merit for some of the missile recorders described. The recorders compared are those which record only – those that cannot fulfill any playback function. Even a fantastically high figure of merit means little if a playback unit is needed also.

Table 2 gives a comparison of figures of merit for missile recorders which do fulfill reproduce functions. It is seen that the Lockheed models AMR-106 through AMR-106 have increasing figures of merit, reflecting the advancement of the "state of the art."

Table 1

MISSILE RECORDERS FOR RECOVERY

MODEL	D_R	f_R	n_T	S/N	V	W	P	FIGURE OF MERIT
MERCURY	5,500	20,000	7	35	640	12	11.2	100,000
DR-25-2	60,000	240	2	35	657	6.9	53	5,070
DR-25A-7	80,000	560	7	35	175	7.4	75	27,500
MTR-362	30,000	60	06	35	70	5	33	21,000
MTR-1400	30,000	899	06	35	560	14	45	12,500

Table 2

MISSILE RECORDERS/REPRODUCERS

MODEL	D_R	f_R	n_T	S/N	V	W	P	FIGURE OF MERIT
MR-100	70,000	300	2	30	600	25	110	337
DR-25-2	30,000	45	6	35	395	11	65	350
MR-400	2,500	1,000	2	35	275	8	35	1,270
MR-40A	2,000	10,000	2	35	250	12	35	6,160
MR-40L	240,000	200	1	30	600	12	35	16,000
600-5	80,000	460	2	30	277	12	57	7,670

THERMOPLASTIC RECORDERS

It appears that thermoplastic recording, which is being developed by General Electric, will have a future in data storage applications where either very high data storage capacity is essential or frequencies above the video range have to be stored. Since the process requires a very well controlled vacuum and extremely high mechanical and optical accuracy, it might not be applicable where weight, volume, and power are factors of prime consideration, or where either limited storage or frequencies below the 5-mc range are needed.

Signal-to-noise ratios on the order of 20 db seem to be optimal and depend on the physical reduction of the light sources used. Ideally, these should be reduced to an infinitesimally small spot in order to increase the signal-to-noise ratio above 20 db. Furthermore, since heating and cooling are contributing to the process, difficulties will arise in limiting ambient temperatures to very narrow margins. The power consumption predicted for airborne recorders in the neighborhood of 200 w seems rather optimistic in the light of the heavy requirements on the use of RF heaters, coolers, vacuum pumps, and associated electronic and mechanical equipment. It does not seem reasonable to expect practical production units to become available during this decade.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to the following companies for their courtesy and cooperation in providing the photographs and descriptions of products presented in this paper, and for the slides used in a previous oral presentation:

Consolidated Electrodynamics Corporation, Pasadena, Calif.
Cook Electric Company, Chicago, Ill.
Jet Propulsion Laboratory, Pasadena, Calif.
Leach Corporation, Compton, Calif.
Speidel Corporation, Providence, R. I.
Lockheed Missiles and Space Division, Palo Alto, Calif.

EFFECTS OF ATMOSPHERIC POLLUTANTS ON ELECTRONIC EQUIPMENT (UNCLASSIFIED)

by

Herbert C. McKee, Manager
Industrial Pollution and Analytical Research
Southwest Research Institute
San Antonio, Texas

INTRODUCTION

Corrosion and other deterioration of military equipment causes serious operational difficulties as well as economic loss. In order to study some of the unusual problems which arise, the Office of Ordnance Research has been conducting work to determine the methods by which these various problems arise. Such studies provide data to serve as a basis for corrective action to increase the reliability of all types of military equipment.

Previous work by Southwest Research Institute under the sponsorship of the Office of Ordnance Research had identified some of the factors which affect the deterioration of Ordnance equipment as a result of exposure to atmospheric humidity. Chief among these factors, other than humidity itself, is the effects of atmospheric contaminants. This effect was illustrated quite graphically by analyzing data obtained in nationwide exposure tests to determine average corrosion rates in different cities. The data were divided by classes based on the average relative humidity of the various cities. Within each relative humidity class, it was quite obvious that the severity of corrosion correlated with the presence or absence of man-made atmospheric pollutants, especially sulfur dioxide. For each relative humidity class, the highest corrosion rates were reported in cities where coal is burned as fuel or where known industrial sources of sulfur dioxide are found. This relationship was not unexpected, since the corrosive effects of sulfur dioxide and other atmospheric pollutants are well known and are documented in the literature.

To obtain preliminary information on some of these problems, the objective of the present project was to collect and analyze samples of deteriorated equipment, to identify any effects which had been caused or modified by the presence of atmospheric contaminants, either particulate or gaseous. This work was to be accomplished in such a way that the effects of humidity could be studied, together with the relationship between humidity and atmospheric contaminants.

Various weapons systems and the control units for these systems were examined at maintenance and rebuild depots, and samples of the system components and of the corrosion deposits and other materials were obtained for chemical analysis. The results of this laboratory analysis then made it possible to determine in many cases what chemical reaction mechanisms had been responsible for the deterioration which occurred.

EXPERIMENTAL RESULTS

Of the various problems which were investigated, those which are of greatest interest to this group involve electronic equipment. Some of these will be described to illustrate the type of information obtained in such a study.

Figure 1 shows an error pulse rectifier used in a control van for a weapon system.

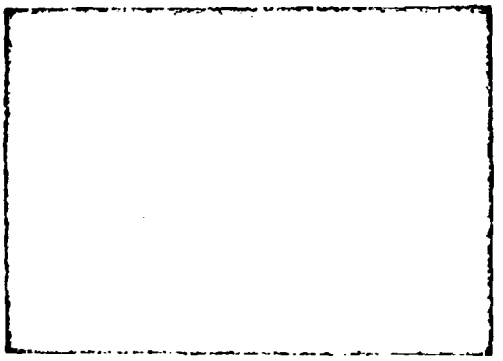


FIGURE 1. ERROR PULSE RECTIFIER, FROM WEAPONS SYSTEM CONTROL VAN

When first examined at the rebuild depot, it was noticed that several of the resistors and other units contained a surface deposit of light colored particles similar in appearance to some types of fungus. A microscopic examination of the chassis and some of the parts also showed very minor pitting which might have been due to chemical attack or might also have been caused by abrasive cleaning, handling or in any of several

other ways. This pitting was not serious enough to impair the efficiency of the unit but the unit was investigated to determine the source of the unusual effects noted.

Typical parts containing this deposit are shown in Figure 2, although the particles themselves are hard to identify due to loss of detail in photographic reproduction. Samples of the particles were removed from the unit and examined

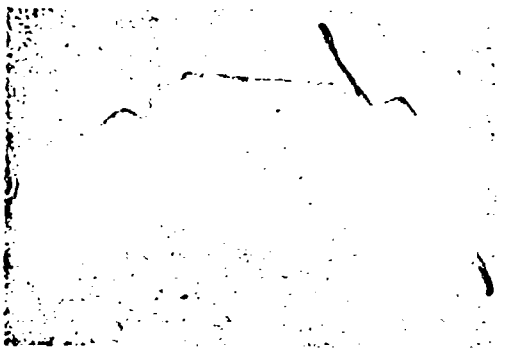


FIGURE 2. TYPICAL PARTS FROM ERROR PULSE RECTIFIER

under a microscope in order to identify the type of deposit. Some of the particles consisted of ordinary dust which had been deposited from the atmosphere. However, a high percentage of the particles were found to be a biological material containing so-called "stone cells." These are a type of vegetable cell structure formed in the growth of vegetation by a thickening and hardening of the material to form a stony or fibrous tissue called sclerenchyma. Frequently these cells pile one on top of another to give the appearance of a pile of stones, which accounts for their name. These cells function chiefly in strengthening and supporting the stem and protecting softer internal tissues or portions of various plants. One of the main sources of stone cells is the hard covering or shell found on several kinds of nuts such as walnuts, almonds, pecans, etc. These shells do not break apart spontaneously and therefore small particles containing these stone cells could not become airborne except by some operation such as grinding or other size reduction.

The identification of stone cells in the deposits presented a rather baffling situation. The most obvious suggestion was that the unit might have been located near a walnut or pecan shelling plant where airborne dust containing small particles of shell material would occur in abundance. However, control vans for military weapons systems usually contain classified equipment, and are customarily located on a missile base many miles from any type of civilian installation.

During a subsequent visit to the rebuild depot it was found that cleaning abrasives composed of ground cellulosic material are used in cleaning some of the equipment. Two different materials are used interchangeably in a gentle cleaning operation similar to sand blasting. One of these materials is ground almond shells, and the second is ground corn cobs. Samples of both were obtained, and it was established by microscopic examination that the ground almond shell material was composed of stone cell structures identical in shape and appearance to the material found on the electronic unit. Thus it seemed certain that the visible deposit originated in the electronics shop rather than at a field operational site as originally supposed. The deposit could have resulted from leaving the unit exposed in the vicinity of the cleaning operation while the ground almond shell material was being used, or through cleaning of this particular unit with the same material.

Since some residue of these materials could be left on practically all of the electronic equipment which was cleaned in this way, further tests were conducted to see if the material might be particularly damaging due to the presence of such residues on electronic parts. By exposing the materials in controlled atmospheres of varying humidity, it was found that they were not hygroscopic; that is, they did not collect water vapor from the atmosphere with the formation of a phase of liquid water around the particles. However, as soon as liquid water collected in the vicinity of the particles by condensation, the particles would then absorb this water and hold it by surface tension much as an ordinary blotter would do. When the humidity was decreased and the surrounding water evaporated into the air, the particles did not become dry as soon as the surrounding surfaces due to the time delay caused by diffusion of water to the surface of the particles from their interior. This time delay was confirmed by a microscopic inspection during drying. Figure 3 shows a particle in the process of drying, with a layer of liquid water visible around the edge of the particle. This photograph was taken several minutes after the surrounding environment had become completely dry. Thus, the particles can serve to collect and hold liquid water as would any other foreign object, especially if porous in nature. However, the particles do not cause condensation or collection of a liquid phase when in contact with humid air below the saturation point.



FIGURE 3. PARTICLE DURING DRYING, SHOWING RETENTION OF LIQUID WATER

Since most types of cellulose-containing biological material contain fungus spores, incubation tests were conducted to see if fungi could be grown from the samples of cleaning material. As expected, it was found that several species of fungus spores occurred in the material, and these spores would start to grow and reproduce if placed in a saturated atmosphere. Figure 4 shows a particle of the ground corncob material as seen through the microscope, with filaments of fungus growth surrounding it. In all some ten



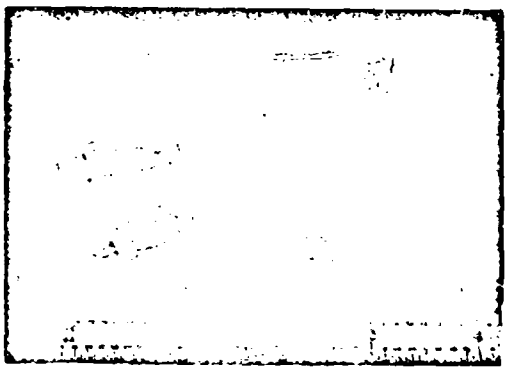
FIGURE 4. PARTICLE OF CLEANING MATERIAL, SHOWING FUNGUS GROWTH

different varieties of fungus were identified, all of them cellulytic fungi which can grow in the presence of moisture and which utilize cellulose as a source of nutrition. Thus, the cleaning material not only provided the fungus spores, but also the cellulose needed for their nutrition.

As a part of their normal growth cycle, cellulytic fungi and in fact all species of fungi produce decomposition products which can be corrosive. Carbohydrates (cellulose) are broken down into simpler sugars and alcohol and during this process the by-products of the filamentous fungi include various organic acids and other potentially corrosive materials.

The practical importance of the occurrence of these cleaning materials on electronic parts is difficult to determine. If they were not present, fungus spores would still be found on the material since various types of fungus spores occur naturally in the atmosphere and would be found on any materials exposed to normal atmosphere. Nutrients would also be available in the form of rubber and plastic insulation, wax-impregnated paper and other materials normally used in electronics equipment. Most of the time these spores are relatively harmless and cause no damage. Only when they are kept in an atmosphere of high humidity (close to saturation) for extended periods of time in the presence of suitable nutrients, do they start to grow and thus produce the potentially corrosive by-products. It is impossible to state whether the use of these cleaning compounds would increase the probability of fungus attack; nevertheless, it does not seem advisable to permit residues to remain which contain both the fungus spores and the material needed for their nutrition.

One special situation was found in which the presence of this cleaning material might cause more trouble than would be the case with a mere surface deposit. Figure 5 shows three resistors taken from the electronic unit previously illustrated. On the left of the picture are two of the units as removed, showing the plastic sleeve which is placed over the resistor unit to prevent contact with the atmosphere. A substantial quantity of the ground almond shell cleaning material can be seen underneath this cover in the annular space between the cover and the resistor. On the right is another unit in which the cylindrical plastic sleeve was cut lengthwise and removed from the resistor. Immediately in front of the resistor is the pile of ground almond shell material which was removed from the annular space. This deposit of cleaning material was cultured and several species of fungi identified. These corresponded to some of the species found in the original material. Spores of these various fungi had been in the resistor unit, remaining dormant in the almond shell material which was trapped between the resistor and the plastic cover.



Left - Normal Units
Right - Cover Removed to Show Deposit Which Was Removed

FIGURE 5. RESISTORS WITH PLASTIC COVER

It is presumed that this material was driven into the annular space between the resistor and the cover by air pressure during the cleaning operation. This represents a potentially more serious threat to the unit than a surface deposit, since the material is held in a very confined space. If the material picked up moisture during an extended period of high humidity it would then dry out very slowly due to being confined. The fungus spores would then be given a longer period of high humidity in which they could grow.

Figure 6 illustrates another unit which was examined, a variable resistor unit contained in a brown plastic case approximately 3 inches in diameter by 1-1/2 inches in length. With this particular item one of the two rotors inside the unit was coated with a white corrosion product. An identical rotor in the bottom compartment of the unit, not shown in the illustration, was unaffected, but was clean with no visible deposits. The white deposit on the top rotor shown in the photograph was removed for chemical analysis, and several determinations were made even though the amount of material was extremely limited. A sample of the metal rotor was also collected by drilling into the metal with an electric drill

The base metal was primarily aluminum plus minor amounts of other alloying materials. The white deposit was composed of various aluminum salts, primarily chloride and ammonium salts. The presence of both chloride and ammonium ions in the deposit suggested the

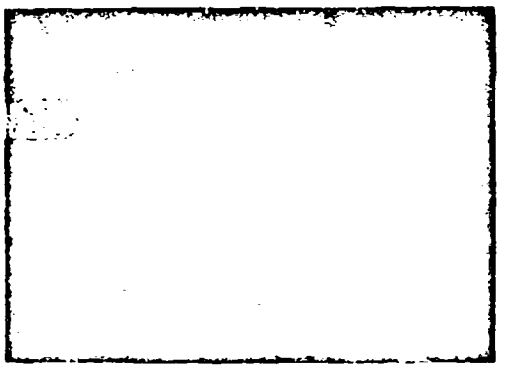


FIGURE 6. VARIABLE RESISTOR UNIT, WITH CORRODED ROTOR

possibility that chlorine or hydrogen chloride as well as ammonia might be involved in the chemical reactions occurring. Some chlorine-containing plastic resins are known to split off hydrogen chloride, which is very corrosive in the presence of moisture. Therefore, a sample of the powdered plastic material was heated with melted sodium, to convert organic chlorides to sodium chloride for analytical tests. However, a test was negative showing that no chlorides were present in the plastic material. A test for ammonia gave positive results indicating that the original plastic material apparently contained nitrogen compounds.

It was also noted that a gasket of rubberized fabric had been used around the top of the resistor to seal the compartment where the corroded rotor was located. Part of this gasket had come off and had been lost but a small portion remained. This small portion was extracted with nitric acid and the extract gave a positive test for chloride indicating that this gasket rather than the base plastic material must have been the source of chloride to assist in the corrosion.

Since both the chloride and ammonia constituents appeared to have been involved in the deterioration, tests were conducted to see if the ammonium content of the base plastic material and the chloride content of the gasket were volatile and therefore could reach the rotor by diffusion through the atmosphere inside of the rotor compartment even though no physical contact occurred between the rotor and the plastic material or the gasket. For this purpose a special apparatus was constructed which is illustrated in Figure 7. A glass tube contained a test reagent solution, while a side arm was provided to

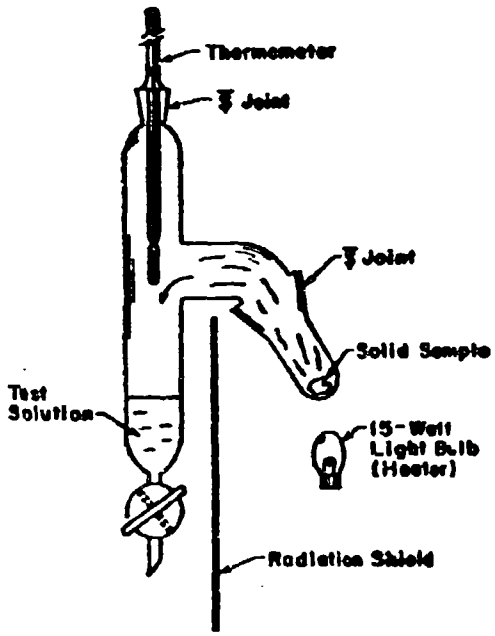


FIGURE 7. APPARATUS TO MEASURE VAPOR FORMATION AND DIFFUSION

contain the material being tested. A small light bulb was placed under the sidearm in order to keep the temperature of the sample a few degrees above the temperature of the reagent solution and thus keep the humidity in the air surrounding the sample slightly below saturation to prevent condensation. The test solution was distilled water containing silver nitrate to test for chlorides, or containing Nessler's reagent to test for ammonia. Any volatile compound evolved by the solid test material would pass by diffusion into the reagent solution, giving the characteristic precipitate for chlorides or yellow color test for ammonia and ammonium compounds.

With the base plastic material, results were positive in the ammonia test and negative in the chloride test. With a small sample of the gasket material, positive results were obtained with chloride and negative results with ammonia. Both the gasket material and the base plastic evolve volatile constituents which evidently affected the corrosion which occurred.

In an attempt to determine whether or not the ammonia and chlorine were both instrumental in causing this particular corrosion problem, small samples of the metal rotor were cut out and suspended in an atmosphere of controlled humidity near but slightly below saturation, together with pieces of the plastic material and the gasket. The sample containing only the gasket showed a small amount of white deposit after an exposure of one week. The sample containing the plastic material showed a similar deposit. The sample containing both the gasket and plastic material showed a white deposit which was somewhat different in crystalline structure as indicated by microscopic examination. The deposit on the latter sample also appeared to etch the base metal underneath, as had occurred with the original rotor, whereas with the other two samples a surface deposit was formed with less evidence of chemical action on the base metal. Additional tests could not be conducted because the small amount of gasket material had been exhausted by this time.

These results indicate very strongly that the corrosion which occurred was the result of the simultaneous action of three factors:

(1) Atmospheric humidity which entered the unit because of the defective gasket.

(2) Ammonia or ammonium compounds evolved by the plastic material.

(3) Chlorine or hydrogen chloride evolved by the gasket material.

As near as can be determined from the small scale tests which were conducted, the absence of any one of these three factors would have prevented this particular type of problem from arising, although humidity plus either of the other two constituents could cause some surface deposit. Based on a small scale test it appeared that etching of the metal rotor would have been less severe without the simultaneous occurrence of all three factors.

SUMMARY

Two specific problems have been described, illustrating the manner in which airborne contaminants, with or without accompanying high humidity, can cause or alter the deterioration of items of electronic equipment. Many more examples of a similar nature could be cited. In some cases precise and detailed detective work is required to determine the chemical and physical reaction mechanisms which are responsible for the deterioration noted. Nevertheless, the type of approach illustrated is frequently successful in identifying the reaction mechanisms involved. This information then leads to a better understanding of the cause of deterioration, which in turn makes it possible to provide for the best and cheapest methods of preventing such deterioration where necessary.

A MATHEMATICAL ANALYSIS OF TRANSIENTS CAUSED BY AGC RESET IN LINE SWITCHING AMPLIFIER AS USED IN AN/MSQ-18 (MISSILE MONITOR) EQUIPMENT

By: John M. Dugan and Richard A. Perry, US Army Air Defense Board

INTRODUCTION: In the use of digital systems for the conveyance of information, the value of error introduced by transients can be of paramount importance. Considered in this paper is the case where one amplifier system is switched between various input lines.

There are two sources from which the transients, seen at the output of the amplifiers, are derived. One source is from within the amplifier itself, and will be termed an internal transient. The other source is from the various input lines, and will be called an external transient. This paper will deal with one transient in particular; however, the results will be perfectly valid for other transients arising from similar sources. The transient considered is associated with AGC build up and causes programming errors in the overall equipment. The original design utilized screening and grid bias potentiometers to balance the circuit and adjust out any transients. This method was unsatisfactory. A modification was then installed in the equipment: that clamped the output of the amplifier to ground for a set time after AGC reset. This method is presently employed but is expensive, large, and only moderately effective. Attempted in this paper, based on a complete analysis of the circuit, will be a new method of transient suppression with accent on comparison to cost, effectiveness, reliability, stability and size. The analysis will be extended to cover initial design considerations in general digital equipment.

TRANSIENT RESPONCE OF AMPLIFIER: The Missile Monitor System employs push-pull amplifiers in the circuits being studied. Figure 1 shows part of the typical circuit and half of the push-pull arrangement. Figures 2 and 3 show the low and high frequency equivalent circuits.

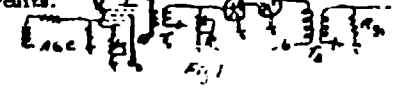


Fig 1

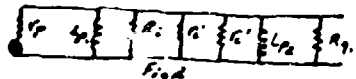


Fig 2

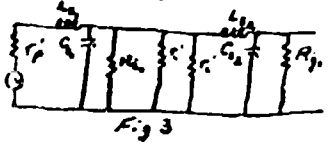


Fig 3

The equations representing this circuit for low frequencies is found to be of the form,

$$Z_{out}(s) = \frac{c_1 s}{a_1 s + a_0} \; ; \; H_{out}(s) = \frac{a_1 s + a_0}{c_2 s + a_0} \qquad Eq\ 1$$

where the constants are real and are functions of the elements only. It can now be shown that the output voltage for a unit step function would be:

$$e_{out} = K_1 \epsilon^{-\alpha_1 t} + C_2$$

where K_1 and α_1 are now functions of the circuit. For the low frequency response of an impulse function, the output voltage is of the nature:

$$e_{out} = K_2 \epsilon^{-\alpha_2 t} + C_2$$

For high frequencies, the circuit takes the form:

$$Z(s) = \frac{a_3 s^3 + a_2 s^2 + a_1 s + a_0}{d_3 s^3 + d_2 s^2 + d_1 s + d_0}$$

$$H(s) = \frac{C_3}{a_3 s^3 + b_2 s^2 + c_1 s^2 + d_2 s + a_0}$$

where again the constants are functions of circuit parameters only. H(s) can be rewritten in the form:

$$H(s) = \frac{C_3}{(s-\beta_1)(s-\beta_2)(s-\beta_3)(s-\beta_4)}$$

for which the inverse transform is of the form:

$$C_3 \left[L_1 \epsilon^{\beta_1 t} + L_2 \epsilon^{\beta_2 t} + L_3 \epsilon^{\beta_3 t} + L_4 \epsilon^{\beta_4 t} \right]$$

$$\text{where } L_n = \prod_{i \neq n} \frac{1}{\beta_n - \beta_i}$$

The output voltage with a unit step input is of the form:

$$\sum_i \frac{K_i}{\beta_i} \epsilon^{\beta_i t} + K_3$$

At $a_1 s \gg b_2$ and $c_1 s \gg d_2$ the circuit moves from the low frequency case to take on mid-frequency characteristics, and the transfer function can be obtained from Equation (1). Also the impedance can be found from Equation (1). Mid-frequency characteristics show themselves in the high frequency case, Equation 2, when $R_{i1} \ll s L_2$, $R_{g1} \ll a_0$, and the terms $s L_{s1}$ and $s L_{s2}$ are negligible. It is seen from these equations that the circuit cannot resonate at low frequencies but will have at least one resonate point at the high frequencies. This point can be shown to be a function of the transformer construction. Also, at low frequencies, the transfer function varies similarly to the transformer characteristic curves.

From standard transformer equivalent circuits it is found that the ratio of amplification of a stage at low frequencies to mid frequencies is equal to:

$$\frac{1}{1 + \frac{r_p}{j\omega L_p}}$$

and for high frequencies to mid frequencies the ratio is:

$$\frac{1/j\omega C_s}{r_p + j\left[\omega L_s - \frac{1}{\omega C_s}\right]}$$

To raise the low frequency half power point it would be necessary to increase r_p or lower L_p. To lower the high frequency half power point c_s or L_s would have to be increased. If L_s is not raised the response falls off slower producing an undesirable effect. It is noticed that a desirable effect can be obtained by lowering L_p for low frequencies and by raising both L_s and c_s in the same ratio for the high frequencies. This gives the desired effects and should be employed in initial design.

DEVELOPMENT OF TRANSIENTS: The particular transient that is being considered here is associated with AGC reset. The problem was traced to the plate circuit of tube V1. Consider the equivalent circuit as shown in Figure 4.

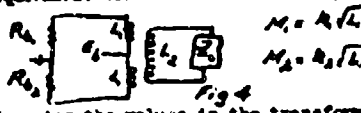


Fig 4

Assuming the values in the transformer to be balanced, and the values of R_b to be functions of e_{gk}. If the current through each tube were always the same, no transient could be observed since the resultant flux due to the D.C. current would be zero. However, considering the case where $R_{b_1} \neq R_{b2}$, both are functions of time, and the plate voltage supply is constant. The equations for the plate current take the form:

$$E_b = (R_{b_1} + sL_1)[I_{b_1} - I_{b_1}(0)] - sM_2[I_{b_2} - I_{b_2}(0)] + sM_1 I_2$$

$$E_b = -sM_2[I_{b_1} - I_{b_1}(0)] + (R_{b_2} + sL_2)[I_{b_2} - I_{b_2}(0)] - sM_1 I_2$$

$$0 = sM_1[I_{b_1} - I_{b_1}(0)] - sM_2[I_{b_2} - I_{b_2}(0)] + (sL_2 + R_2)I_2$$

where

$$I_{b_1}(0) = 0 \; ; \; I_{b_2}(0) = \frac{E_b}{R'_{b_2}}$$

Assume the amplifier is balanced at one value of E_c. Then $R_{b1} = R'_{b2}$

$$\Delta I = \frac{E_b}{s R'_{b_2}} \left[\frac{R_{b_2} R_2 (sL_2 + R_2) - R_b R_2 (sL_2 + R_2)}{\nabla} \right]$$

where $R_1 = R_b'' + R_b' R_{b_1}$

$$R_2 = R_b'' + R_b' R_{b_2}$$

$$\nabla = [s^2 L_1 (R_{b_1} + R_{b_2}) + R_{b_1} R_{b_2}](sL_2 + R_2)$$

Which gives for a solution:

$$\Delta I = I_o [1 - e^{-\theta t}]$$

where

$$\theta = \frac{R_{b_1} R_{b_2}}{2L_1 (R_{b_1} + R_{b_2})} \; ; \; I_o = \frac{E_b (R_{b_2} - R_{b_1})}{R_{b_2} R_{b_1}}$$

Which will produce a voltage in the secondary of:

$$e_2 = -\theta I_o e^{-\theta t} = \frac{E_b (R_{b_2} - R_{b_1})}{2L_1 (R_{b_1} + R_{b_2})} e^{-\theta t}$$

The plate current can be approximated in the region of operation by:

$$I_p = I_{po} e^{\alpha (E_c + \mu)} + A E_g + C$$

By adjusting the bias and screen grid voltage, the stage can be balanced at one particular operating point. However it is noticed that the value of R_b is dependent on α , which normally varies as much as 300% between tubes.

It was found by a series of tests, that the normal variation from the average value of R_p at any particular voltage can be expected to be less than 25%. This permits, however, a sizeable transient to be developed even when the amplifier is balanced at one particular voltage. This is brought about by the fact that during AGC reset, the bias voltage is clamped to -6 volts. It then builds up rapidly to around -11 volts.

Possible means to eliminate this transient would be to add a second stage in this section with R-C coupling between the present tube and the added one. The added tube could then drive the transformer, and by the correct choice of values for the R-C network, the effects of the transient could be eliminated.

It is noticed in this analysis that an excellent rule to follow in using this type of equipment is to never permit sudden changes of D.C. current in any transformer. This was appreciated when a clamp was attempted during the initial design of a modification to eliminate the transient. It was noticed at that time that the clamp caused a sudden change of D.C. current flowing in T_2, and caused a larger transient than the one being clamped out.

Comparison of calculated values from these equations and actual values photographed are displayed at the end of the paper.

DESIGN OF A FILTER: During the analysis it was noticed that the bandwidth on the Missile Monitor system is rather broad, and it was recommended that a filter be designed and placed in the system. A well designed filter would not only decrease the effects of the previously studied transient, but would also block external transients from reaching the data handling portion of the equipment.

There exists a certain amount of freedom in the placement of the filter; however, it would be advantageous to place it as near the output as possible to clean the lines of internal transients. The exact placement of the filter designed for this paper would be in the output lines to one battery where 600 ohms is seen in both directions where the lines are broken. This facilitates the design of the filter and makes the construction of it much simpler. However, if a filter were to be installed permanently in the equipment, it should be placed earlier in the circuit where one filter would filter all the outputs.

The design is rather straight forward, using two band pass filters in parallel, one for 600 and one for 1500 cycles per second. Care must be taken to have the characteristic impedances total 600 ohms. This can be done with two symmetric filters since the impedance increases rapidly away from the pass band. Figure 5 shows the proposed filter.

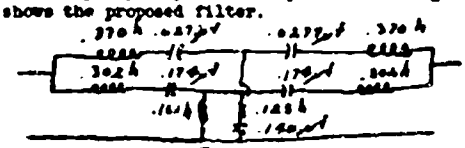


Fig 5

At the present time the filter has not been built and tried due to immediate unavailability of parts and time. It is planned though to build and test this filter at an early date. In this particular filter, the phase relationship between the 1500 cycle and 600 cycle frequencies would be unaltered. The value of the 1500 cycles passing through the 600 cycle filter would be down 1.5 nepers, and the value of the 600 cycles passing through the 1500 cycle filters would be down 3.1 nepers. This should provide no trouble due to leakage of one frequency through the other frequency's filter.

COMPARISON: The recommendation of adding

a second tube to the AGC stage is the best method of eliminating the trouble in this particular system. However, in present models this would necessitate an excessive amount of labor. It is hoped that the designed filter will be adequate for these systems, and if so, will provide a much cheaper and more reliable method of eliminating transients. At the same time the designed filter will suppress external transients which remain a problem with the present clamp modification in use. The size of the filter may be slightly larger than the present clamp.

The addition of a second tube for new systems should be attempted. The flip-flops, saturable reactor, and logic gates now in use could be excluded thus eliminating a great deal of the cost. It would provide a higher degree of reliability since an amplifier fails less often than does a flip-flop or diodes used in gates. The size, due to the added tubes, would be slightly larger, but the wiring and construction would be simplified. Overall, it is considered to be much more effective and reliable than the existing method of clamping the line.

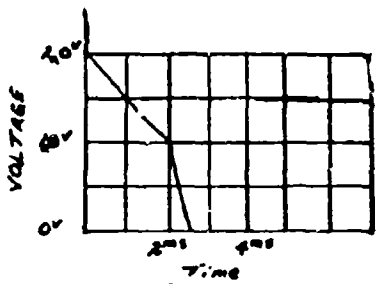
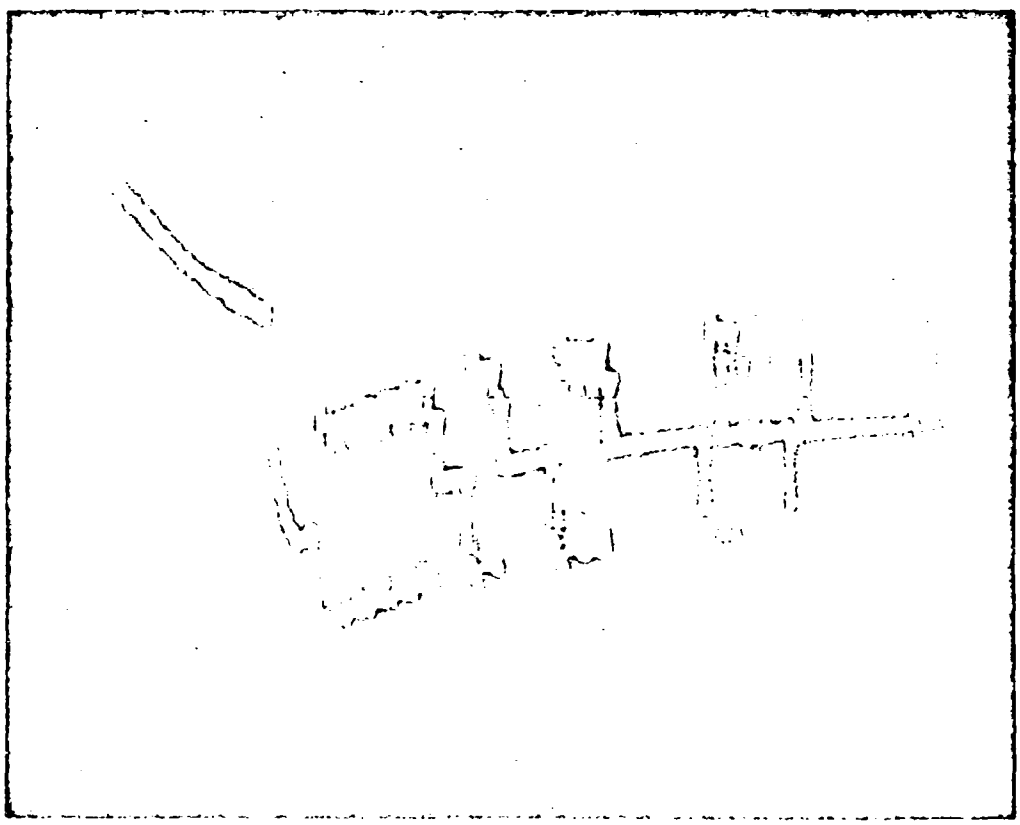


Fig 6
Calculated Transient

Fig 7
Actual Transient
Voltage base 0.5v/div vs. Time 0.5ms/div



AN ANALOGUE COMPUTER FOR SEPARATING EVOKED PHYSIOLOGICAL POTENTIALS FROM BACKGROUND NOISE

By: Walter J. Kropfl, Robert L. Robinson
John C. Arrington, Donald I. Tapes

Walter Reed Army Institute of Research

Electrical activity produced in the living organism is a major concern of electrophysiology. In general there are two methods of recording activity. One is by using microelectrodes to record the electrical activity of single cells. This type of recording produces pulses which have durations in the order of one millisecond and frequencies of more than 500 cycles per second. This method requires surgical operation on the subject and is therefore not suitable for normal human subjects.

The second method, the one in which we are most interested in this paper, uses gross electrodes. Gross electrodes may be applied to the subject's skin. For example, electrodes on the surface of the scalp produce potentials known as electroencephalogram (EEG). The EEG, for the most part, is an ongoing activity having an irregular waveform with a peak to peak amplitude ranging up to 100 uV. If the subject observes an intense flickering light synchronized waves may be evoked in the EEG. Typically, however, these evoked potentials are much smaller than the activity which is not associated with the stimulus. Thus, they can not be easily investigated using conventional recording procedures.

A response to flickering light may also be recorded from the eye. It is obtained by using a gross electrode imbeded in a contact lens. The lens is filled with a conducting solution that makes contact with the surface area of the eye. These recordings thus obtained are called electroretinograms. If very bright stimuli are used the ERG is readily discerned. As the luminance of the stimulus light is reduced it becomes weaker until it, too, is hidden by background noise.

Since gross electrodes record the activity of large numbers of cells, an extremely complex waveform may be obtained. Figure 1 shows a typical example of an ERG response from the eye. The waveform changes when the light is flickering at different rates. At the 4/second rate the waves marked 1 and 2 are obtained. It has been shown that the wave marked 1 is produced by the photopic or day mechanism of the retina, and that marked 2 is produced by the scotopic or night vision system. At the higher rate of 20 flashes per second only the photopic component is obtained. The electroretinograms shown in the figure were found with bright stimuli. At lower luminances they would be hidden by the background noise. At moderate luminances the variable nature of evoked responses makes quantitative analysis of the data in its original form extremely unreliable, particularly when they are partially marked by noise.

A system for detecting these evoked responses in the presence of an adverse signal to noise ratio has been developed, and will be described in this paper. It combines features of devices which have been described by Dawson 1, Barlow 2, and Rosenblith et al. 3. Its advantages include its simplicity, flexibility and low cost.

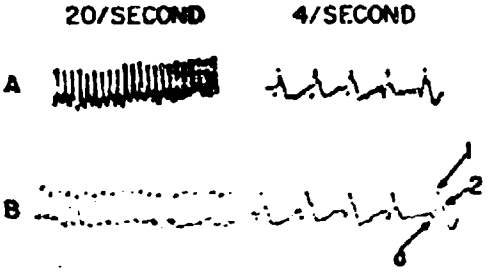


Figure 1

Figure 2 shows a block diagram of the system. Both the electroretinogram and occipital (EEG) potentials are recorded on an FM tape recorder. Each time a stimulus light flashes a photocell triggers a pulse generator which is also recorded on tape. This stimulus pulse is used to synchronize the control unit. It triggers a chain of ten single shot multivibrators. Nine of these one shots control the duration and closing time of the relays. The action is such that after an initial time delay each member in the chain of relays closes and opens in sequence. The remaining one shot delays the computing time by a preset amount.

The path of the signals through the relay contacts and computer is shown in more detail in Fig. 3. The response signal passes through a limiter to the relay contacts, K_1, K_2, K_3 etc. which feed it to a bank of operational amplifiers connected as integrators. They store and sum the activity during the time intervals T_2-T_1, T_3-T_2---$T_{N-1}-T_N$ as shown. The output of each integrator is proportional to the sum of the areas lying under successive response curves during fixed time intervals following each stimulus. Potentiometers at the output of the integrators are used to compensate for differences in the pull in and drop out times of the relays. A second set of relay contacts k_1, k_2 etc. closes at the same time and in the same sequence as the input relays. This set of relay contacts samples the output of each integrator in turn, and delivers the signal to an oscilloscope.

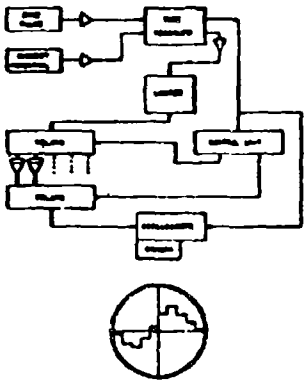


Figure 2

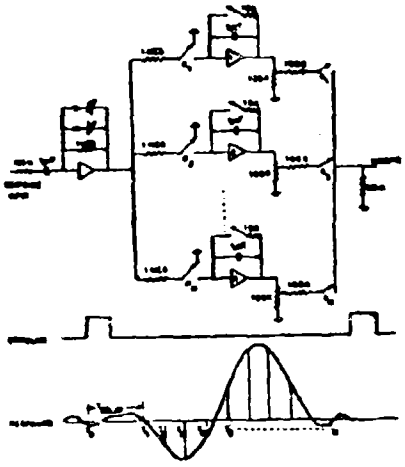


Figure 3

Whenever a subject winks, the rotation of his eye produces a large potential several times the amplitude of the normal ongoing activity. The limiter at the input of the computer reduces the error produced by such artifacts.

The times during which the relays are closed are under the control of the experimenter. If only a rough approximation of the response is desired the nine relay closings are spaced to cover the entire response interval. If a more detailed picture is required, the relays are set for brief time samples. Only a portion of the response waveform will be reconstructed. However, the tape may be played back a second time with the relays delayed to a later part of the response wave. Several playings with the relays delayed at increasing intervals make it possible to reconstruct the whole wave with a high degree of resolution. The smallest sampling segment is 2 milliseconds long and is limited by the relays. If the tape is played back at a slower rate than that at which it was recorded, it is possible to make effective time samples shorter than 2 milliseconds. On the other hand, if the tape is played back faster it is possible to obtain a number of samples of the waveform in a short time but with a sacrifice of resolution.

Before discussing some of the results which illustrate the value of this computer, a brief description of the associated experimental apparatus is in order. The experimental apparatus consists of a conventional visual stimulator and recording system as well as the actual computer. A sketch of this system is shown in Figure 4.

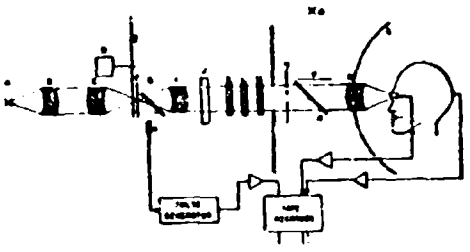


Figure 4

Light from a tungsten ribbon filament (A) is focused on the plane of a motor driven rotating sectored disk (E). The rays which pass through the sector are collimated and transmitted through a series of interchangeable neutral density and color filters (K, L, M) before being presented to the subject in Maxwellian view. A camera shutter (J) is used to remove the stimulus from the subject's view. A series of interchangeable stops (N) behind the final lens enables the experimenter to control stimulus diameter. A surrounding screen (S) and an auxiliary optical system (O, P, Q) make it possible to hold the subject's eye at a fixed level of adaptation. The adaptation field nearly fills the retina including the area where test flashes are presented. A series of fixation points are used so that stimuli may be located at various retinal positions.

The electrical response of the eye, the electroretinogram, has been used as a response for checking the performance of the computer. This was done because it is a comparatively reliable response about which much is already known. It can be seen above the noise level when bright stimuli are used. Thus, it was possible to follow it downward below the noise with progressively diminishing stimulus luminances. The electroretinogram was a good response to use for checking the resolution of the system. Most of the electroretinograms investigated so far have been elicited by flickering orange and blue light at the rates

of 18 and 5 flashes/sec. The orange and blue lights were selected to favor photopic and scotopic response activity respectively. Photopic activity is related to the cone receptors of the retina whereas scotopic activity is related to rod activity.

Figure 5 shows photographs obtained with orange stimuli and with the relays gated so as to sample the whole response waveform.

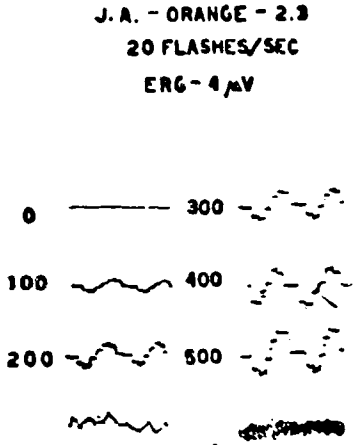


Figure 5

The bottom tracing in the left column shows the activity as it is picked up at the electrodes. The noise, however, is so large that no response can be detected visually. The picture at the bottom of the right hand column shows a number of sweeps superimposed. There is still no evidence of a response. The remaining photographs show the output of the computer as increasing numbers are added. The retinal response can be detected readily even when a relatively small number of responses are added, and it stands out more and more clearly as the number is increased.

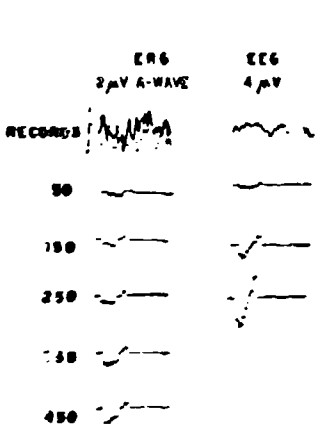


Figure 6

Figure 6 shows photographs of the growth of a blue response. The ERG is shown in the left column. As in the previous figures, it grows and becomes more clearly developed with an increasing number of additions. The right column shows samples of summated brain potentials. Although they have not yet received detailed analysis, they are readily detected by the computer.

The computer has been used in experiments directed towards stray light problems. Typically, the electroretinogram is seen only in response to strong stimulation. Glaringly bright flashes produce stray light which stimulates the retinal areas outside of the actual image. It is difficult to obtain any evidence of localized response activity. For example, when a bright red photopic stimulus is imaged on the fovea which contains many cone receptors, the response is no greater than when the stimulus is imaged on a part of the retina containing very few cones. If stimuli of low luminance are delivered to the retina when it is already flooded with a steady adaptation light, it might be supposed that the effectiveness of stray light would be reduced and that there would be more evidence for localization. The computer makes it possible to record the extremely small responses which are obtained when these conditions are fulfilled. Preliminary data have been obtained for stimuli of various luminances and field sizes. In addition, small stimuli have been spotted on different retinal positions to determine whether the sensitivity of different regions differs when the effectiveness of stray light has been reduced.

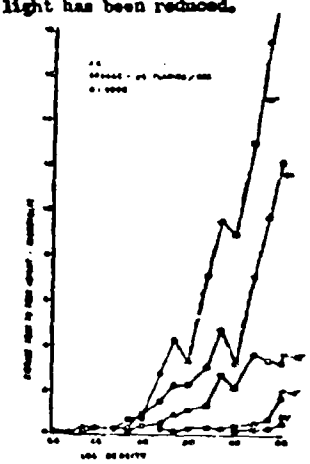


Figure 7

Figure 7 shows a sample of luminance curve in which the peak to peak height of the ERG is plotted versus the amount of neutral filtering inserted in the stimulus path for orange stimuli of several diameters. Each point on the curve represents a single summation. It took several experimental sessions to obtain the data presented in this graph, and some of the irregularities in the curve can probably be attributed to day to day variabilities. For all diameters response decreases systematically with decreasing intensity.

The slope of the luminance curve changes with stimulus diameter. This curve illustrates that responses whose average peak to peak amplitude is less than 1 uV can be elicited by averaging techniques.

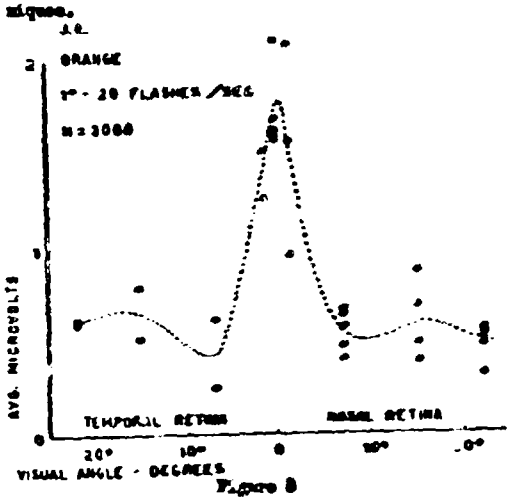


Figure 8

Figure 8 shows the height of the response obtained at different retinal positions for a constant stimulus luminance. It is clear that the sensitivity of the retina is not uniform. This result cannot be obtained using bright stimuli and conventional recording procedures. Thus, even early data indicate that simple computers can be of great value in electrophysiological experimentation.

References

1. Lawson, G. D.: EEG & Clin. Neurophysiol., 1954, 6:65.

2. Communications Biophysics Group of Research Laboratory of Electronics and W. H. Siebert. Massachusetts Institute of Technology [illegible] Cambridge, [illegible].

3. Barlow, J. S.: EEG of Clin. Neurophysiol., 1954, 6:1.

PRECISION FREQUENCY MEASUREMENT OF DOPPLER SATELLITE SIGNALS

by: William A. Dunn, Ballistic Measurements Laboratory,
Ballistic Research Laboratories, Aberdeen Proving Ground,
Maryland

ABSTRACT

The Ballistic Research Laboratories have conducted laboratory tests on the frequency uncertainty expected in measuring the frequency of a radio doppler signal employing the data gathering equipment of the DOPLOC (Doppler Phase LOCk) satellite tracking station. A brief description of the test layout system will be given including the timing system, tracking filter, and the recorder-readout equipment. The purpose for the tests will be discussed and the test setup described. The calculated theoretical frequency measuring capabilities of the system will be discussed and compared with empirical results obtained from the laboratory tests. The test results---RMS frequency error as a function of signal-to-noise ratio and tracking filter bandwidth---will be discussed. An indication of propagation effects will be shown through a discussion of the analysis of field data from actual satellite passes.

INTRODUCTION

The Ballistic Research Laboratories have been engaged in tracking satellites since the early days of Sputnik I in October 1957. The DOPLOC (Doppler Phase LOCk) satellite tracking system soon came into being. The rf signal received at the DOPLOC complex is heterodyned to audio frequencies by mixing with a signal generated at the station by use of high stability frequency standards so that all of the data measurements are made at audio frequencies. The facilities at the DOPLOC complex make it possible to obtain either a direct or indirect measure of frequency. The direct measurement is obtained by simply counting the cycles for a given time interval---usually one second--- while the indirect frequency measurement is obtained by measuring the elapsed time (period) for a given number of Doppler cycles and then converting that time measurement to frequency. The frequency counters have the capability of $\pm$ 1 cycle which yields a counting precision of $\pm \sqrt{2}$ cycles per second when both ends of the interval are considered. In making the period measurement a 100 kc signal from a frequency standard is fed to the [illegible] counter which also has the capability, with [illegible] cycle or, the equivalent in time, $\pm$ 10 microseconds. The indirect frequency measuring method yields the same precision as the direct frequency measuring method when the period, Δt, is equal to f/100,000 where f is the frequency of the signal to be measured. If the period is greater than f/100,000 the precision is enhanced, i.e., counting 1000 cycles of a 2000 cps signal gives a reading precision of $\sqrt{2}/2$ cps while counting 2000 cycles of a 2000 cps signal gives a reading precision of $\sqrt{2}/50$ cps.

The analysis of data from satellite pass indicated, in some cases, an RMS random scatter of as much as 5 cycles. Even though there were many explanations for this scatter, an increasing concern developed in some quarters over the actual capabilities of the phase locked tracking filter which is the heart of the DOPLOC system. A series of laboratory tests were made on data simulating that from actual satellite passes. The results of these tests yielded a random scatter in the data essentially equal to the precision of measurement thus absolving the tracking filter of any blame but it also had the effect of creating a keen interest in attempting to make an empirical determination of the capabilities of the tracking filter to either confirm or refute its calculated theoretical capabilities. It was readily apparent that to accomplish this task required considerably greater reading precision than was presently available. It was as a result of this realization that the precision of the period measurement was increased by a factor of 100 by feeding a 10 mc signal to the time counter as a replacement for the 100 kc signal. The tests, to be discussed later, involved.

DESCRIPTION OF DOPLOC STATION COMPUTER

A block diagram of the basic Doppler satellite tracking station is shown in Figure 1. The diagram shows the system as set up for a satellite transmitting at 108 mc; however, any frequency can readily be accommodated by merely changing one local oscillator frequency and one rf amplifier circuit in the preamplifier. The tracking station at Aberdeen is presently equipped with four such channels of equipment giving it the capability of tracking on more than one frequency for satellites carrying a multi-frequency transmitter.

A brief description of some of the individual components of the system and their application follows:

1. Timing System

The timing system is composed of a Borg 100 kc frequency standard, the output of which is fed to a time code generator which converts the 100 kc to one per second time pulses. The one per second pulses are fed to a clock and synchronized with a time signal received from WWV. All of the reader-recorder equipment is controlled by the clock, thus synchronization of all measurement and readout data is assured.

2. BRL Tracking Filter[1]

The tracking filter presently in use is the Model 4 phase-locked filter available from the Interstate Electronics Corporation. A very basic block diagram of the tracking filter system, extracted from Reference 1, shown in Figure 2.

---

1. V. W. Richard, BRL Memorandum Report No. 1175, October 1958, "DOPLOC Tracking Filter".

It is an electronic band pass filter and the filtering action is obtained by the use of a voltage controlled oscillator that is correlated or phase locked to the input signal frequency. Tracking is accomplished with an electronic servo system designed to make the voltage controlled oscillator follow the variations of frequency and phase of the input signal. Several filter bandwidths are available---0.5, 1, 2.5, 10, 25, and 50 cps. The lower the bandwidth the greater S/N improvement of the filter is obtained; however, the higher the bandwidth the greater rate of change of frequency of the input signal can be tolerated. Some compromise must be made on filter bandwidth depending on the characteristics of the signal. The bandwidth, however, can be manually adjusted and can be changed during actual tracking without losing lock or introducing excessive phase transients into the data.

3. Computing Digital Indicator and 10 mc Counter

The computing digital indicator is a Dymec 2500. It is a versatile piece of equipment which in its application in the DOPLOC system serves a dual role---as a counter and also as a gate for the 10 mc counter which is a Hewlett Packard 524B. If a direct frequency measurement is to be made the Dymec is controlled by the clock one per second time pulses and counts the number of Doppler cycles between two successive time pulses and then feeds its output to the digital printer which is a Hewlett Packard modified 560-A. If an indirect frequency measurement (period) is desired the number of Doppler cycles to be sampled is set in the Dymec and, on command from the one per second clock pulse, opens a gate to the Hewlett Packard 10 mc counter on the first Doppler cycle crossover following the command and closes the gate on the last Doppler cycle of the number selected. The output of the 10 mc counter is then fed to the Hewlett Packard digital printer.

The Hewlett Packard 524 B 10 mc counter derives its basic frequency from the Borg 100 kc frequency standard in the timing system and internally multiplies this up by a factor of 100 yielding 10 mc. It should be noted that, in making the period measurement, the number of cycles to be selected by the Dymec 2500 is completely arbitrary and can be set at any convenient value.

4. Hewlett Packard Digital Printer

The Hewlett Packard modified 560-A digital printer is utilized to furnish printed output data as a function of time on paper tape. The printer has an eleven digit capacity. When a direct frequency measurement is made six of the eleven digits are used to record real time in hours, minutes, and seconds and the remaining five digits record frequency data. When the indirect frequency measurement is made and it is desirable to record the time period to the nearest tenth microsecond it becomes necessary to utilize seven digits for the data measurement which leaves only four available for real time---minutes and seconds---and the station operator writes in the hours.

DISCUSSION OF THE THEORETICAL CAPABILITIES OF THE TRACKING FILTER

Reference 1 states that the signal-to-noise improvement by the tracking filter is that of the theoretical ratio of input noise bandwidth to tracking filter bandwidth. Expressed as S/N voltage ratios,

$$\text{Output } S/N = \text{Input } S/N \sqrt{\frac{B_{input}}{B_{t.f.}}} \qquad (1)$$

where $B_{t.f.}$ is the tracking filter bandwidth and B_{input} is the noise bandwidth of the input signal. The theoretical signal-to-noise improvement by the tracking filter is shown in Figure 3. It relates signal-to-noise improvement to the tracking filter bandwidth for several input noise bandwidths. By taking the output from the voltage controlled oscillator the output peak amplitude of the filter is constant and independent of the input signal level. The filter effectively converts the amplitude noise components of the input to phase noise or jitter at the output. It is shown in reference 1 that the RMS output phase noise in radians is $1/\sqrt{2}$ of the RMS input noise-to-signal ratio multiplied by the square root of the ratio of tracking filter bandwidth to input noise bandwidth, i.e.

$$\theta_o = \frac{1}{\sqrt{2}} \frac{N}{S} \sqrt{\frac{B_{t.f.}}{B_{input}}} \qquad (2)$$

where θ_o = RMS output phase jitter in radians. Solving equation (1) for the input N/S ratio and substituting in equation (2) we have

$$\theta_{o_{RMS}} = \frac{1}{\sqrt{2}\left(\frac{S}{N}\right)_{output}} \qquad (3)$$

Since this phase jitter can be expected at both ends of the measuring interval we find it necessary to multiply equation (3) by the $\sqrt{2}$ and, for convenience, we convert from radians to cycles by dividing by 2π. The equation then becomes

$$\sigma_c = \frac{1}{2\pi \left(S/N\right)_{output}} \qquad (4)$$

where σ_c is the RMS phase jitter at the output of the tracking filter expressed in cycles. Phase jitter is equal to frequency jitter in cycles per second when the integrating period is equal to one second. In order to make equation (4) applicable to any integrating interval and to express the results as frequency uncertainty the equation is rewritten as

$$\sigma_{cps} = \frac{1}{2\pi(\Delta t)\left(\frac{S}{N}\right)_{output}} \qquad (5)$$

where σ_{cps} is the RMS frequency error in cycles per second and (S/N)output is expressed in vol-

tage ratios*. We call 1/2π the theoretical cision constant for the tracking filter. Fig 4 shows the theoretical calculated RMS frequen error and the theoretical output signal-to-noise ratio in db as a function of input signal-to-noise ratio for three tracking filter bandwidths---1, 10, and 50 cps---and a 20 kc input noise bandwidth.

DESCRIPTION OF FREQUENCY MEASUREMENT TESTS

In conducting the actual frequency measurement tests some minor variations were made in the DOPLOC instrumentation system. A block diagram of the test setup is shown in Figure 5. The basic signal was derived from a Rohde and Schwarz Model XUA EH&&&462 frequency synthesizer, fed directly into the counter-readout system, and its frequency recorded. The same signal is mixed with noise generated by a General Radio noise generator and the resultant noisy signal passed through the tracking filter, then through the counter-readout system, and its frequency recorded. Since the two counter-recorder systems are synchronized in time any drift in the original frequency will be of no consequence in the analysis of the data. The frequency measurement of the input data was considered as a standard and the residuals and an RMS error were computed for the tracking filter output. All of the test runs were made with a 20 kc input noise bandwidth. Input frequencies of 2, 4, 6, 8, and 10 kc were used at two tracking filter bandwidths---10 and 50 cps. The measurement mode was an indirect frequency (period) measurement and the number of Doppler cycles selected to sample was chosen so that the integrating period was one second for the sake of simplicity in the computations. A summary of the test runs made is given in the following table:

Frequency	Input S/N(db)	Tracking Filter Bandwidth
2,4,6,8,10 kc	0	10, 50 cps
"	-6	"
"	-12	"
"	-18	"
"	-24	"

A total of forty tests were made and the duration of each test was between three and four minutes.

*Woodward and Gebor give a similar general relationship for frequency uncertainty, $\Delta F = 1/t_i(S/N)$, where ΔF is the frequency uncertainty, t_i is the integrating period, and S/N is expressed in voltage ratios. Radar designers have their own similar equation, $\Delta t = 1/BW(S/N)$, where Δt is related to range jitter, BW is the bandwidth and is related to integrating time, and S/N is expressed in voltage ratios.

EMPIRICAL RESULTS AND COMPARISON WITH THEORETICAL CALCULATED RESULTS

Figure 6 shows the frequency scatter, or the difference between the input frequency and the output frequency, as a function of time, for a case where the input frequency was 10 kc, the input signal-to-noise ratio was -18 db, and the tracking filter bandwidth was 10 cps. Peak excursions of 0.16 cps are observed and for this particular test the RMS frequency error is 0.052 cps.

Figure 7 shows a comparison of the empirical results and the results expected based on theoretical calculations for all of the test runs made with a 10 cps tracking filter bandwidth. RMS frequency error in cycles per second and output signal-to-noise ratio in db are related to the input signal-to-noise ratio in db. Curve A shows the empirical results while curve B shows the results based on theoretical calculations. It is noted that the RMS frequency error is higher by about a factor of two than those expected from theoretical calculations and that the empirical output signal-to-noise ratio is about 6 db below the theoretical values.

Figure 8 shows a similar comparison for all tests run with a 50 cps tracking filter bandwidth. As expected, the accuracy of the frequency measurement is lessened with a 50 cps bandwidth but what is even more significant is that the empirical result in RMS frequency error is, again, larger by about a factor of two than those expected from theoretical calculations.

DISCUSSION OF ERRORS IN ACTUAL FIELD DATA

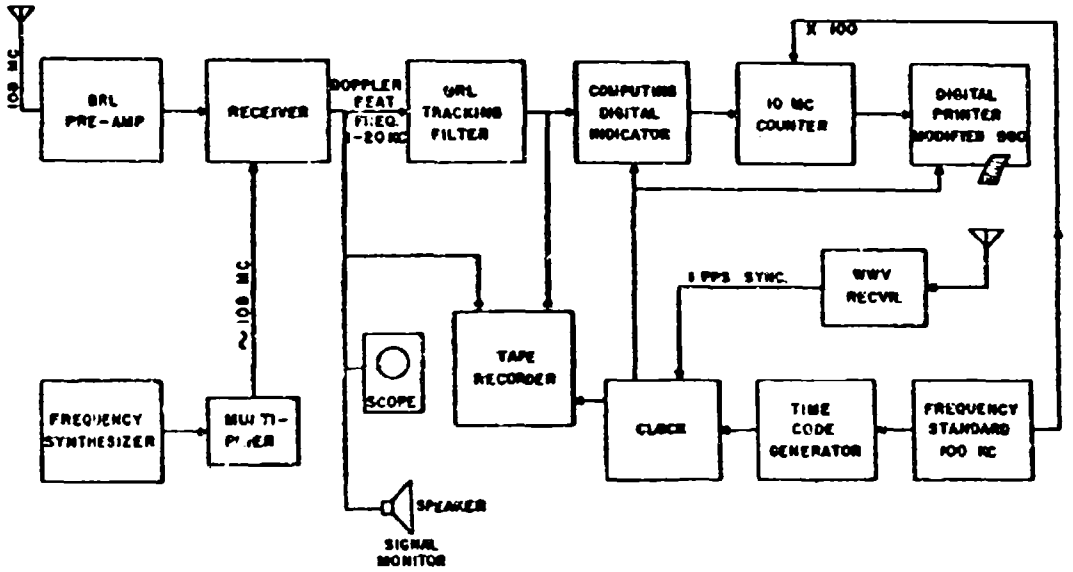
It was mentioned in the introductory remarks that the RMS random scatter in data received from actual satellite passes had been observed to be as great as 3 cycles. Although no intensive study has been made of actual satellite data for the purpose of analyzing the random scatter, the general observation has been made that the scatter is less for satellites that pass nearly over the receiving station. In addition, it appears in general, that the scatter is less for satellites which have nearly circular orbits. It appears that the conclusion can be drawn that, to a large degree, the random scatter in actual recorded satellite data is a function of propagation phenomena. It should be noted that random scatter could be introduced to some extent as a result of satellite spin or, to a somewhat less degree, as a result of instability in the satellite transmitter.

CONCLUSIONS

All of the tests and the results presented were obtained with tracking filter bandwidths of 10 and 50 cps for convenience in performing the tests and some slight gain in simplicity of calculations. However, in practice, a very large percentage of data from actual satellite passes has

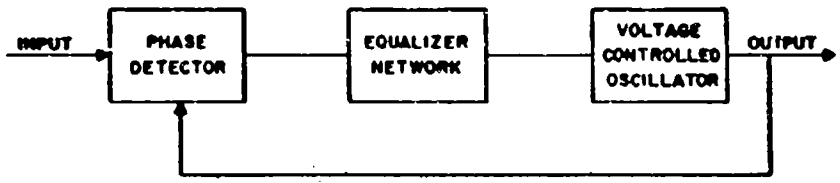
been gathered using a 5 cps tracking filter bandwidth. This enhances the accuracy of the frequency measurement by about 25 percent over that demonstrated for the 10 cps bandwidth.

The fact that the empirical results show a 2 to 1 relationship in accuracy with those obtained based on theoretical calculations is neither surprising nor discouraging. It is felt, instead, that the results are most encouraging. One does have a tendency to become suspicious, however, when they obtain a lack of comparison by almost exactly a factor of two and the present case was no exception. For the purpose of these tests the theory has been accepted at face value. The empirical results, regardless of their lack of agreement with theoretical calculated results, speak for themselves and a high degree of precision has been exhibited in making frequency measurements of noisy Doppler signals.



108 MC. DOPLOC STATION
BASIC DOPPLER SATELLITE TRACKING
STATION USING BRL TRACKING FILTER

FIGURE 1



BASIC TRACKING FILTER SYSTEM

FIGURE 2

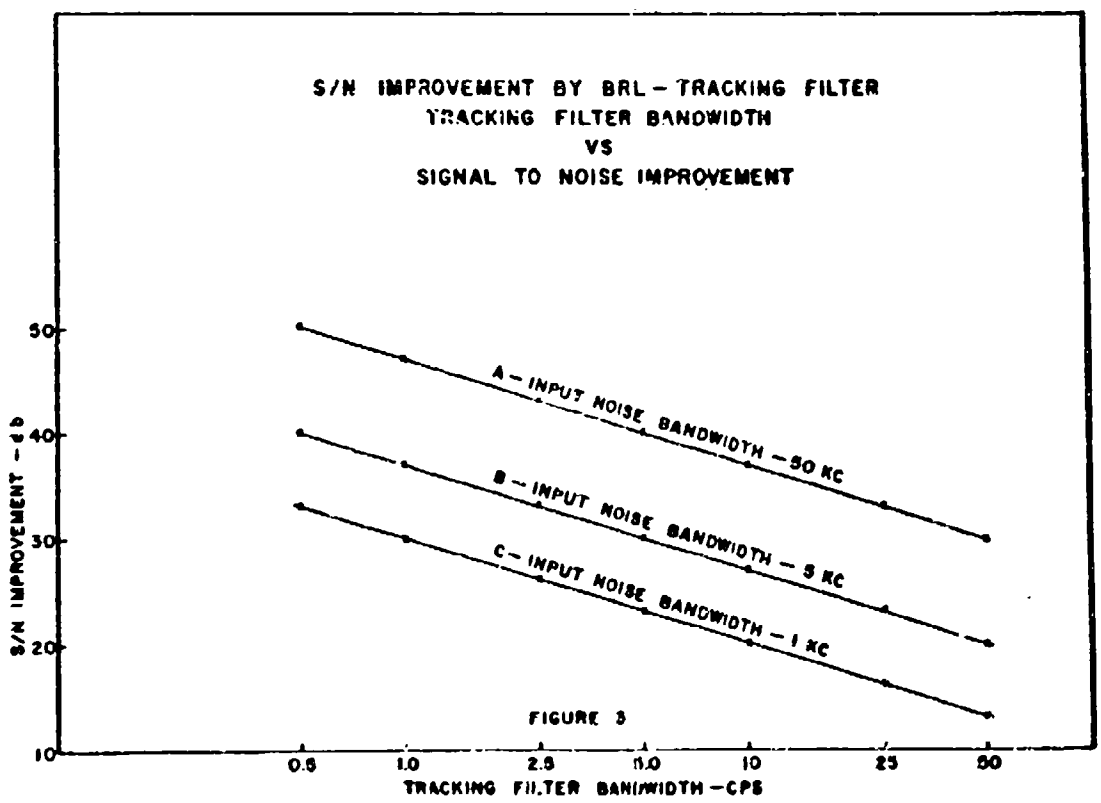


FIGURE 3

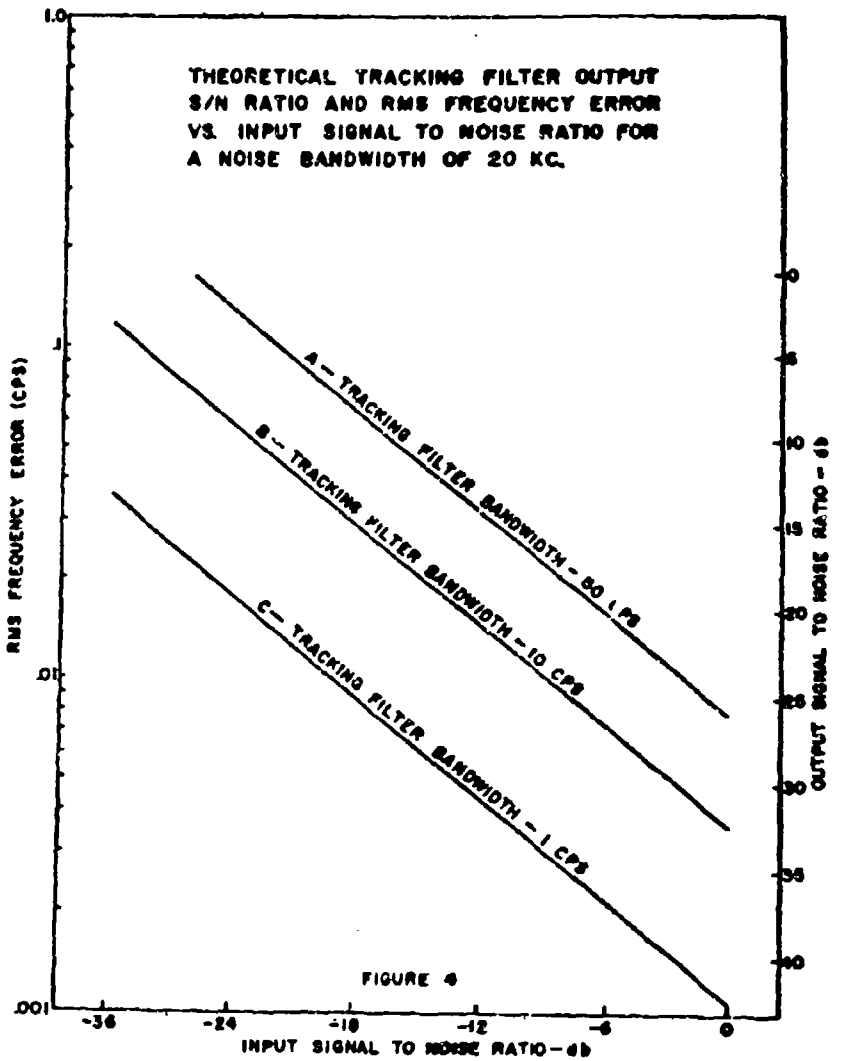
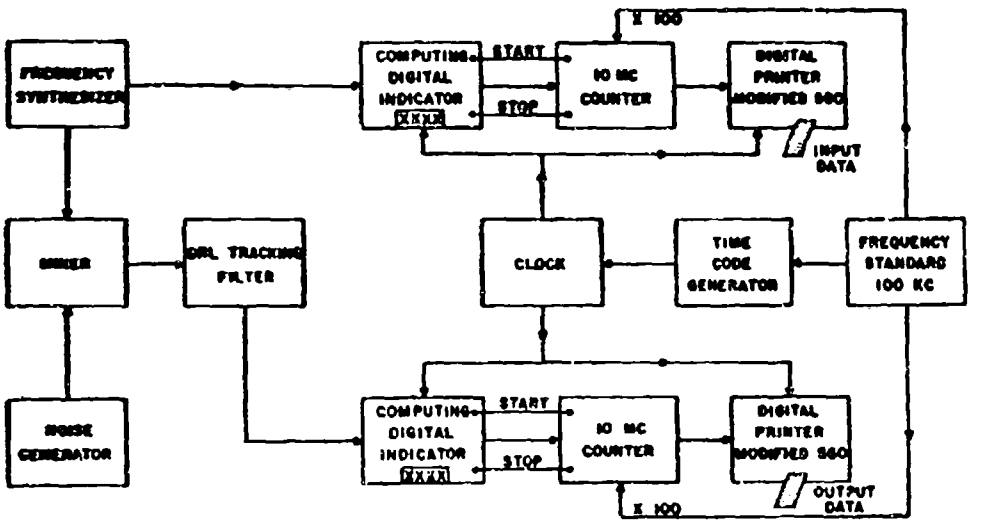

THEORETICAL TRACKING FILTER OUTPUT S/N RATIO AND RMS FREQUENCY ERROR VS. INPUT SIGNAL TO NOISE RATIO FOR A NOISE BANDWIDTH OF 20 KC.
A — TRACKING FILTER BANDWIDTH = 50 CPS
B — TRACKING FILTER BANDWIDTH = 10 CPS
C — TRACKING FILTER BANDWIDTH = 1 CPS
RMS FREQUENCY ERROR (CPS)
1.0
.1
.01
.001
OUTPUT SIGNAL TO NOISE RATIO — db
0
5
10
15
20
25
30
35
40
-36
-24
-18
-12
-6
0
INPUT SIGNAL TO NOISE RATIO — db

FIGURE 4



DOPLOC STATION EQUIPMENT AS USED
IN FREQUENCY MEASUREMENT TEST

FIGURE 5

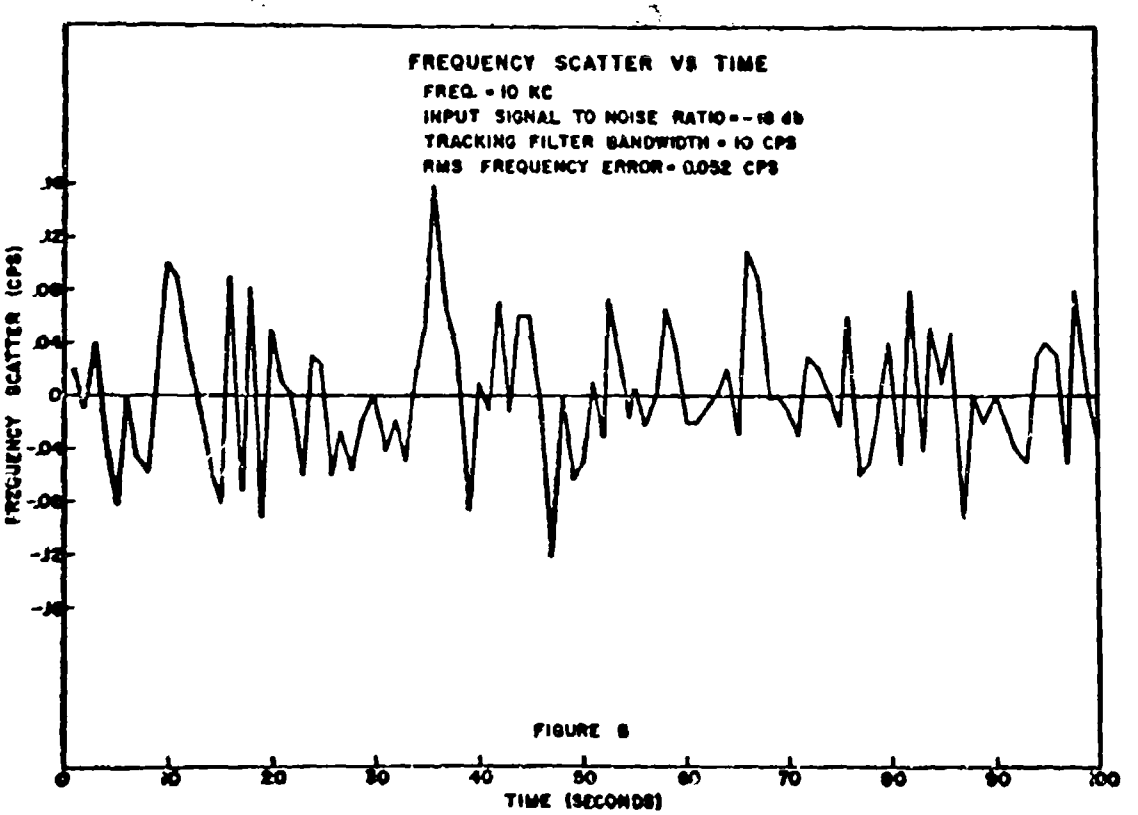


FIGURE 6

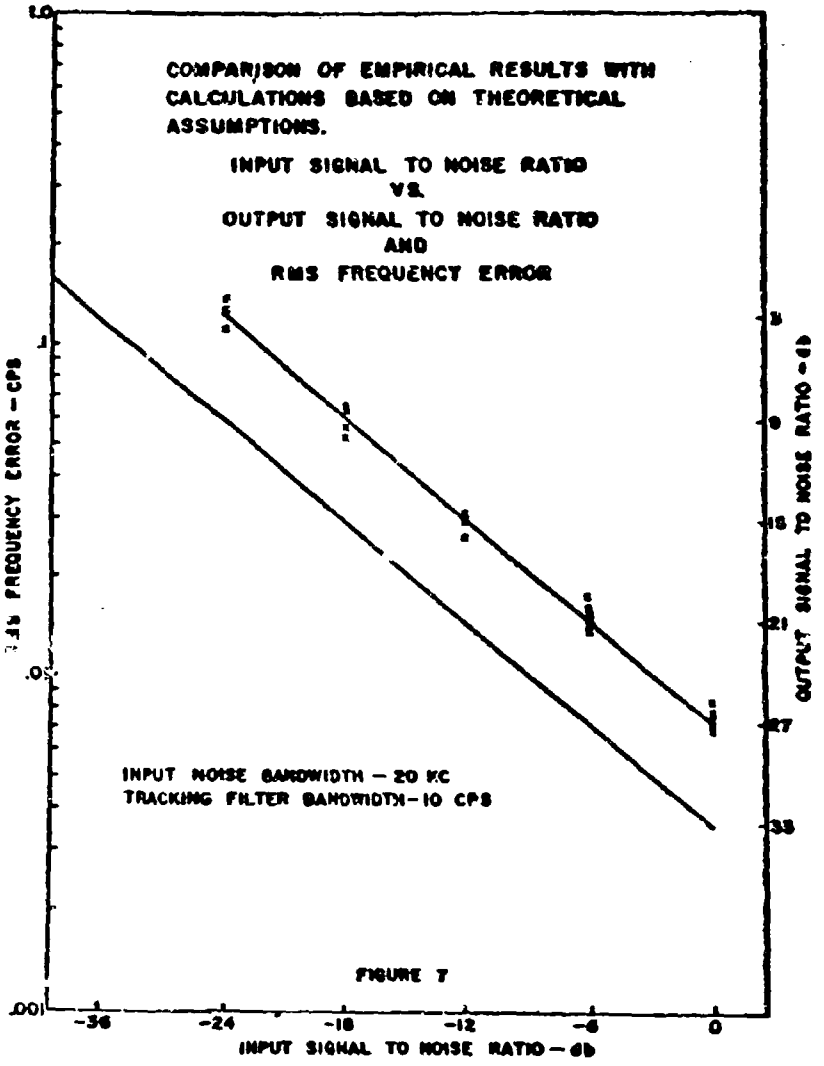

COMPARISON OF EMPIRICAL RESULTS WITH CALCULATIONS BASED ON THEORETICAL ASSUMPTIONS.
INPUT SIGNAL TO NOISE RATIO
VS.
OUTPUT SIGNAL TO NOISE RATIO
AND
RMS FREQUENCY ERROR
INPUT NOISE BANDWIDTH — 20 KC
TRACKING FILTER BANDWIDTH — 10 CPS
RMS FREQUENCY ERROR — CPS
OUTPUT SIGNAL TO NOISE RATIO — db
INPUT SIGNAL TO NOISE RATIO — db
1.0
.1
.01
-36
-24
-18
-12
-6
0
18
21
27
33

FIGURE 7

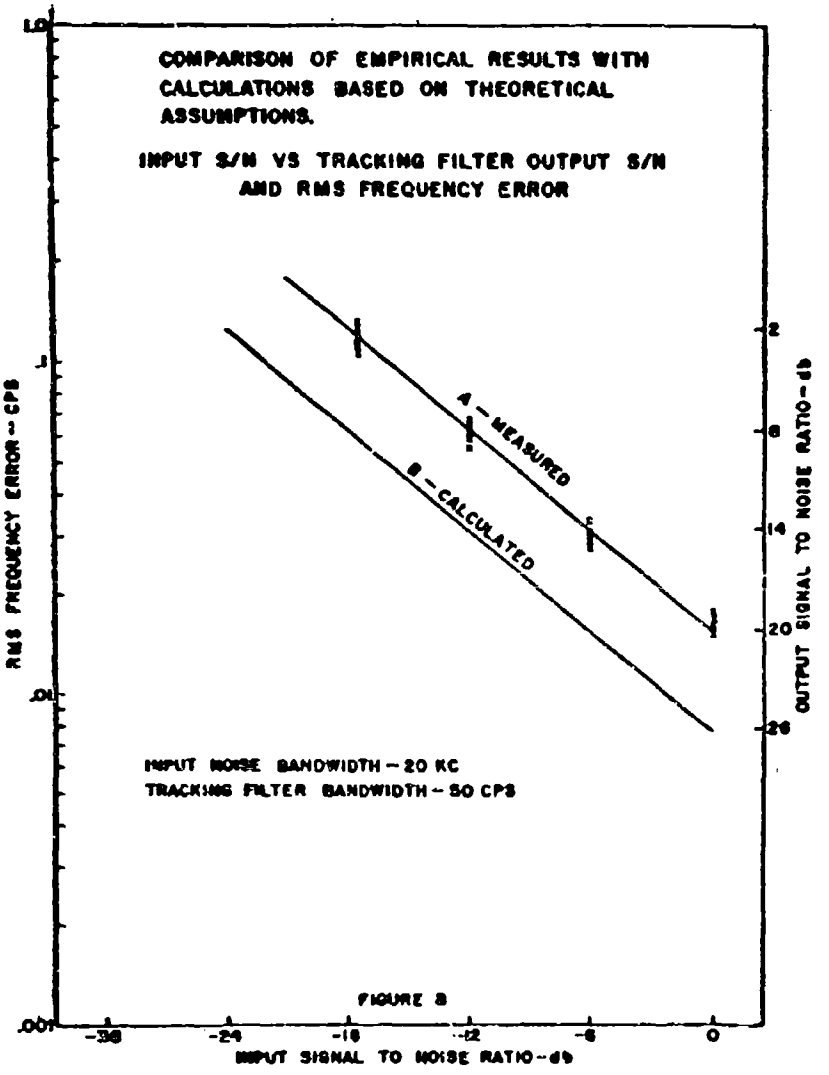

COMPARISON OF EMPIRICAL RESULTS WITH CALCULATIONS BASED ON THEORETICAL ASSUMPTIONS.
INPUT S/N VS TRACKING FILTER OUTPUT S/N AND RMS FREQUENCY ERROR
A – MEASURED
B – CALCULATED
INPUT NOISE BANDWIDTH – 20 KC
TRACKING FILTER BANDWIDTH – 50 CPS
RMS FREQUENCY ERROR – CPS
OUTPUT SIGNAL TO NOISE RATIO – db
INPUT SIGNAL TO NOISE RATIO – db
1.0
.1
.01
.001
2
8
14
20
26
-30
-24
-18
-12
-6
0

FIGURE 8

OUTPUT SIGNAL-TO-NOISE CHARACTERISTICS OF CORRELATORS

By: Bruce R. Wayo, Heavy Military Electronics Dept., General Electric Co., Syracuse, N. Y.
and, David K. Cheng, Electrical Engineering Dept., Syracuse University, Syracuse, N. Y.

ABSTRACT

This paper presents a detailed analysis of the output signal-to-noise ratio for a correlator consisting of a multiplier followed by an RC averager. Both the signal and the noise are assumed to be band-limited functions with Gaussian amplitude distribution; but no restrictions are imposed on either the input signal-to-noise ratio or the product of the bandwidth and the time constant of the averager. A family of curves are plotted relating output and input signal-to-noise ratios for values of RCΔf ranging from 0.1 to 100. The dependence of output signal-to-noise characteristics on the relative values of bandwidth and center frequency is discussed.

INTRODUCTION

One of the most important measures of the efficacy of correlators as signal detectors is the output signal-to-noise ratio. Because the averaging period in a practical correlator can not be infinite, a fluctuating component about the desired correlation function appears at the output as noise. The mean-square output signal-to-noise ratio can be defined as the ratio of the square of the desired correlation function and the mean-square noise when there is no relative delay at the inputs. In general, the output signal-to-noise ratio of a correlator depends upon the input signal-to-noise ratio, the center frequency and bandwidth of the input signal and noise, and the time constant of the averager in a complicated way.

Much work has been done on the analytical calculation of the output signal-to-noise ratio of correlators.[1][2][3] The work by Faran and Hills[3] was particularly comprehensive. However, with a view to simplifying analysis, the assumption is usually made that the time constant (RC) of the averager is very large in comparison with the reciprocal of the bandwidth ($1/\Delta f$) of the signal. With this assumption the dependence of the output signal-to-noise ratio on the center frequency is lost. Also lacking is the information as to how large the product RCΔf must be in order that the calculated values from simplified analyses would represent good approximations.

The restriction on the value of the product RCΔf is removed in this paper, and a general expression for the output signal-to-noise ratio of a correlator consisting of a multiplier followed by an R-C averager is obtained. The quantity $f_0/\Delta f$ appears in the expression as a parameter, and the dependence of the output signal-to-noise ratio on the center frequency is placed in evidence. The conditions under which the signal-to-noise ratio at the output will be higher than that at the input are also shown.

OUTPUT S/N IN TERMS OF CORRELATION FUNCTIONS

The block diagram of a correlator with two inputs is shown in Fig. 1. The waveforms at the inputs to the multiplier are represented by

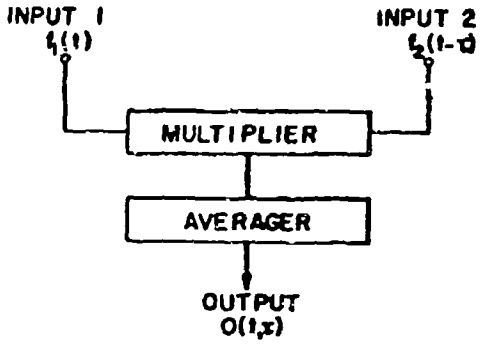


Fig. 1-- Block diagram of a correlator

$f_1(t)$ and $f_2(t-\tau)$, where τ is the time delay. If $h(t)$ is the impulse response of the averager, then the output, $O(t,\tau)$, of the correlator can be written as a convolution integral as follows:

$$O(t,\tau) = \int_0^\infty f_1(t-t')\, f_2(t-t'-\tau)h(t')dt' \qquad (1)$$

The mean-square value of $O(t,\tau)$ is

$$\overline{O^2(t,\tau)} = \int_0^\infty\int_0^\infty \overline{f_1(t-t')f_2(t-t'-\tau)f_1(t-t'')f_2(t-t''-\tau)}\, h(t')h(t'')dt'\,dt'' \qquad (2)$$

where the overhead bars indicate time averages. For inputs with a Gaussian distribution the averaged quadruple product in the integrand can be expressed in terms of autocorrection functions R_{11} and R_{22}, and cross-correlation functions R_{12} and R_{21}.[3] Equation (2) becomes

$$\overline{O^2(t,\tau)} = \int_0^\infty\int_0^\infty \Big[R_{12}^2(\tau)+R_{11}(t''-t')\ R_{22}(t''-t') + R_{12}(t''-t'+\tau)R_{21}(t''-t'+\tau)\Big]h(t')h(t'')dt'dt'' \qquad (3)$$

If a transformed weighting function $W(\ )$ is defined such that

$$W(x) = \int_{-\infty}^{\infty} h(t)h(t+x)dt \qquad (4)$$

Eq. (3) is reducible to the following form:

$$\overline{o^2(t,\tau)} = R_{12}^2(\tau) + \int_{-\infty}^{\infty} W(x)\left[R_{11}(x)R_{22}(x) + R_{12}(x+\tau)\,R_{21}(x-\tau)\right]dx \qquad (5)$$

Equation (5) expresses the mean-square output of the correlator as the sum of the square of the desired correlation function (the signal), $R_{12}^2(\tau)$, and a mean-square noise term. The output signal-to-noise ratio of the correlator for an arbitrary time delay τ is then

$$\left(\frac{S}{N}\right)_o = \frac{R_{12}^2(\tau)}{\int_{-\infty}^{\infty} W(x)\left[R_{11}(x)R_{22}(x) + R_{12}(x+\tau)R_{21}(x-\tau)\right]dx} \qquad (6)$$

Equation (6) is a general expression for Gaussian inputs; no restrictions have yet been imposed on the frequency spectra of the inputs or on the response characteristics of the averager. With τ set to zero (maximum mean-square output signal), Eq. (6) reduces to

$$\left(\frac{S}{N}\right)_o = \frac{R_{12}^2(0)}{\int_{-\infty}^{\infty} W(x)\left[R_{11}(x)R_{22}(x) + R_{12}(x)R_{21}(x)\right]dx} \qquad (7)$$

CHARACTERISTICS OF R-C AVERAGER

Consider the simple R-C averaging network in Fig. 2. The transfer function $H(j\omega)$ is



Fig. 2 - A simple R-C averager

readily seen to be

$$H(j\omega) = \frac{1}{1 + j\omega RC} \qquad (8)$$

and the impulse response is

$$h(t) = \frac{1}{RC}\,e^{-t/RC}, \quad t > 0$$
$$= 0, \quad t < 0 \qquad (9)$$

Using Eq. (4), the transformed weighting function $W(x)$ is easily found.

$$W(x) = \frac{1}{2RC}\,e^{-|x|/RC} \qquad (10)$$

Graphical plots of $h(t)$ and $W(x)$ are shown in Fig. 3.

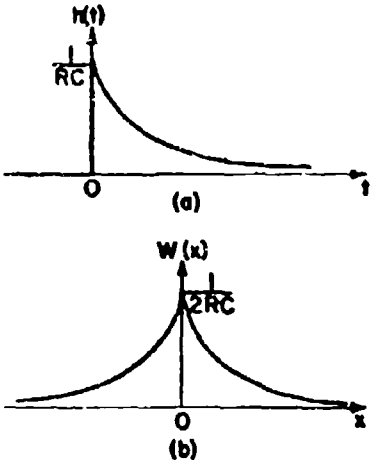


Fig. 3 - $h(t)$ and $W(x)$ of simple R-C averager

OUTPUT S/N FOR BAND-LIMITED SIGNAL AND NOISE

Let it be assumed that:

1. Both the signal and the noise are Gaussian and possess a rectangular spectrum of half-width Δf centered at frequence f_o.
2. The signals at the two inputs are coherent and each has a mean-square value S_1.
3. The noises at the two inputs are uncorrelated and each has a mean-square value N_1.

Under these conditions the correlation functions become

$$R_{11}(x)=R_{22}(x)=(S_1+N_1)\,\frac{\sin 2\pi\Delta fx}{2\pi\Delta fx}\cos 2\pi f_o x \quad (11)$$

$$R_{12}(x)=R_{21}(x)=S_1\,\frac{\sin 2\pi\Delta fx}{2\pi\Delta fx}\cos 2\pi f_o x \quad (12)$$

Substituting Eqs. (10), (11), and (12) in Eq. (7), one obtains

$$\left(\frac{S}{N}\right)_o = \frac{2RCS_1^2}{\left[(S_1+N_1)^2 + S_1^2\right]D} \quad (13)$$

where

$$D = \int_{-\infty}^{\infty} e^{-|x|/RC}\left(\frac{\sin 2\pi\Delta fx}{2\pi\Delta fx}\cos 2\pi f_o x\right)^2 dx \quad (14)$$

The evaluation of the integral in Eq. (14) is tedious, but it can be done with the aid of available tables.[4] The final result is given below:

$$D = \frac{1}{2\pi\Delta f}\left\{\tan^{-1}(4\pi RC\Delta f)\right.$$

$$+\frac{1}{2}\tan^{-1}\frac{2(4\pi RC\Delta f)\left[1+(4\pi RCf_o)^2\right]}{1-(4\pi RC\Delta f)^2\left[1+(4\pi RCf_o)^2\right]+(4\pi RCf_o)^2\left[2+(4\pi RCf_o)^2\right]}$$

$$-\frac{f_o}{2\Delta f}\tan^{-1}\frac{2(4\pi RCf_o)(4\pi RC\Delta f)^2}{1+(4\pi RC\Delta f)^2\left[1-(4\pi RCf_o)^2\right]+(4\pi RCf_o)^2\left[2+(4\pi RCf_o)^2\right]}$$

$$\left.-\frac{1}{4(4\pi RC\Delta f)}\ln\frac{1+(4\pi RC)^2(f_o+\Delta f)^2}{1+(4\pi RC)^2(f_o-\Delta f)^2}\right\} \quad (15)$$

Calling

$$\Delta\theta = 4\pi RC\Delta f \quad (16)$$

$$\theta_o = 4\pi RCf_o \quad (17)$$

and substituting Eq. (15) in Eq. (13), one finally has

$$\left(\frac{S}{N}\right)_o = \frac{(S_1/N_1)^2}{(1+S_1/N_1)^2+(S_1/N_1)^2}$$

$$\times\left\{\frac{\Delta\theta}{\left[\tan^{-1}\Delta\theta+\frac{1}{2}\tan^{-1}\frac{2\Delta\theta(1+\theta_o^2)}{1-(\Delta\theta)^2(1+\theta_o^2)+\theta_o^2(2+\theta_o^2)}\right.}\right.$$

$$-\frac{\theta_o}{2\Delta\theta}\tan^{-1}\frac{2\theta_o(\Delta\theta)^2}{1+(\Delta\theta)^2(1-\theta_o^2)+\theta_o^2(2+\theta_o^2)}$$

$$\left.\left.-\frac{1}{4\Delta\theta}\ln\frac{1+(\theta_o+\Delta\theta)^2}{1+(\theta_o-\Delta\theta)^2}\right]\right\} \quad (18)$$

The output signal-to-noise ratio $(S/N)_o$ is plotted versus the ratio of the mean-square signal to mean-square noise (S_1/N_1) in Fig. 4 using the product $RC\Delta f$ as a parameter for the cases when $f_o = 100\,\Delta f$ ($\theta_o = 100\,\Delta\theta$) and when $f_o = \Delta f$ ($\theta_o = \Delta\theta$). It is seen that $(S/N)_o$ increases almost as the square of (S_1/N_1) when S_1 is about 8 db below the level of N_1. As the input noise diminishes relative to the signal, $(S/N)_o$ increases more slowly with (S_1/N_1); and for random Gaussian signals the output signal-to-noise ratio of the correlator approaches a saturation level even when the input noise diminishes to zero. The saturation $(S/N)_o$ when there is no noise at the input of course depends upon the relative values of the signal bandwidth and the passband of the averager, hence on the product $RC\Delta f$.

The curves in Fig. 4 also show that the output signal-to-noise characteristics of a correlator depend upon the center frequency f_o unless the product $RC\Delta f$ is greater than or equal to 1. The unity-slope line is a line of demarcation. Points lying above this line represent conditions under which the output signal-to-noise ratio is higher than the input signal-to-noise ratio. This demonstrates clearly the noise-reduction capabilities of correlators.

Two special cases are of particular interest. They are: (1) $\Delta\theta \gg 1$ and $\theta_o \gg \Delta\theta$, and (2) $\Delta\theta \gg 1$ and $\theta_o = \Delta\theta$. These two cases are examined below. It is noted that the condition $\Delta\theta \gg 1$ is approximately equivalent to the requirement $RC\Delta f > 1$.

(1) $\Delta\theta \gg 1$ and $\theta_o \gg \Delta\theta$ ($RC\Delta f > 1$ and $f_o \gg \Delta f$).

Under these conditions the second and third terms in the denominator of the expression inside the curved brackets of Eq. (18) approximately cancel, the fourth term vanishes, and the first term is the only effective term which is

$$\tan^{-1} \Delta\theta \approx \pi/2.$$

Equation (18) reduces to

$$\left(\frac{S}{N}\right)_o = 8RC\Delta f \frac{(S_i/N_i)^2}{(1+S_i/N_i)^2 + (S_i/N_i)^2} \qquad (19)$$

which becomes

$$\left(\frac{S}{N}\right)_o \approx 8RC\Delta f \left(\frac{S_i}{N_i}\right)^2, \text{ when } \left(\frac{S_i}{N_i}\right)^2 \ll 1 \qquad (19a)$$

and

$$\left(\frac{S}{N}\right)_o \approx 4RC\Delta f, \text{ when } \left(\frac{S_i}{N_i}\right)^2 \gg 1 \qquad (19b)$$

(2) $\Delta\theta \gg 1$ and $\theta_o = \Delta\theta$ ($RC\Delta f > 1$ and $f_o = \Delta f$).

Similar to case (1) above, the first term approximately equals $\pi/2$, the second and third terms cancel, but the fourth term no longer vanishes.

$$\frac{1}{4\Delta\theta} \ln \frac{1+(\theta_o+\Delta\theta)^2}{1+(\theta_o-\Delta\theta)^2} \approx \frac{1}{4\Delta\theta} \ln (2\Delta\theta)^2 = \frac{1}{\Delta\theta} \ln \sqrt{2\Delta\theta}$$

$$= \frac{1}{4\pi RC\Delta f} \ln 2\sqrt{2\pi RC\Delta f}$$

Equation (18) becomes

$$\left(\frac{S}{N}\right)_o = \left\{\frac{8RC\Delta f}{1 - \frac{1}{2\pi^2 RC\Delta f} \ln 2\sqrt{2\pi RC\Delta f}}\right\} \frac{(S_i/N_i)^2}{(1+S_i/N_i)^2+(S_i/N_i)^2} \qquad (20)$$

In comparison with Eq. (19), the coefficient in Eq. (20) is $1.002(8RC\Delta f)$ when $RC\Delta f = 100$, and $1.082(8RC\Delta f)$ when $RC\Delta f = 1$. When $RC\Delta f < 1$, neither Eq. (19) nor Eq. (20) represents a good approximation for the general expression in Eq. (18).

DISCUSSION

A general formula expressing the output signal-to-noise ratio of a correlator in terms of the input signal-to-noise ratio, the center frequency and bandwidth of the input signal and noise, and the time constant of the averager has been obtained. Both the signal and the noise are assumed to be Gaussian random functions and to have a band-limited rectangular frequency spectrum. However, the formula holds approximately true also for small arbitrary signals in the presence of Gaussian noise of high spectral density. The output signal-to-noise ratio increases almost as the square of the input signal-to-noise ratio when the input signal is weak compared with noise, and it approaches a saturation level as the noise at the input diminishes toward zero. The dependence of the output signal-to-noise ratio upon the center frequency f_o and the product $RC\Delta f$ is examined. It is found that as long as $RC\Delta f$ is greater than 1 approximate formulas appeared in the literature give good results and center frequency f_o has little effect. However, the effect of f_o on output signal-to-noise ratio can not be neglected when the product $RC\Delta f$ is small.

REFERENCES

(1) R. M. Fano, "Signal-to-Noise Ratio in Correlation Detectors," MIT Research Laboratory of Electronics, T.R. No. 186; February 19, 1951.

(2) W. B. Davenport, Jr., "Correlator Errors due to Finite Observation Intervals," MIT Research Laboratory of Electronics, T.R. No. 191; March 8, 1951.

(3) J. J. Faran, Jr., and R. Hills, Jr., "Correlators for Signal Reception," Acoustics Research Laboratory, Harvard University, Technical Memorandum No. 27; September 15, 1952.

(4) A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, "Tables of Integral Transforms," McGraw-Hill Book Company, New York, N. Y. (1954).

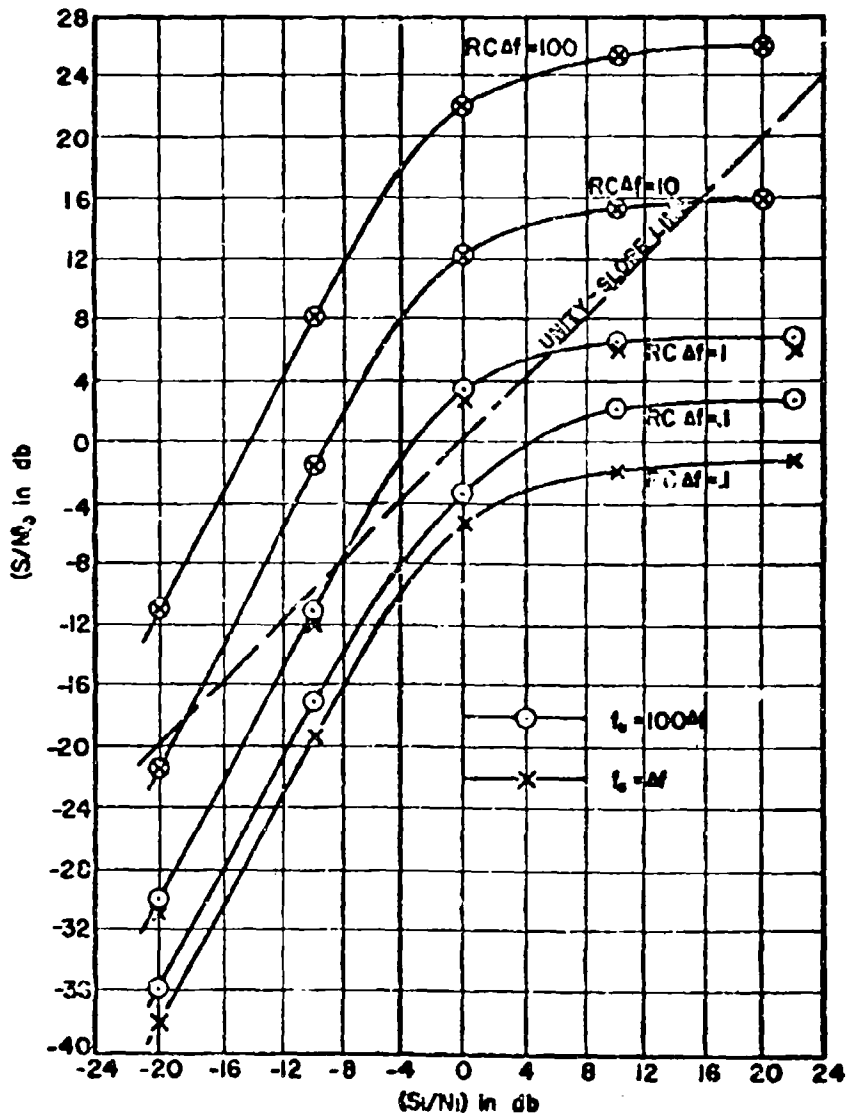


FIG. 4 - OUTPUT SIGNAL-TO-NOISE CHARACTERISTICS AS FUNCTIONS OF Si/Ni.

MILITARY APPLICATIONS FOR SPEECH COMPRESSION TECHNIQUES

By: A. J. Strassman and K. C. Stockhoff
Communications Division
Hughes Aircraft Company, Los Angeles, California

Requirements

Despite the rapid progress being made in the field of providing digital methods of communications, it must be realized that there will be continued, and even increased, usage of voice communications throughout the military complex. Even the male species of the human race will not forego the opportunity to "pass the word," nor is it conceivable that any General will be willing to give up his prerogative for "reading off" a recalcitrant Colonel.

It has been stated that military requirements for communications facilities of all types are increasing at a rapid rate, due primarily to the missile and satellite applications which were nonexistent as little as ten years ago, and which are now demanding, and will require, more and more of the available spectrum to accomplish their assigned tasks. Coupled with the requirement for additional communications facilities is the fact that the spectrum can handle only a limited amount of traffic because of bandwidth restrictions. Despite the inroads made by digital techniques in the communications field, it has been estimated that at least 50 percent of all communications will still be by voice.

Since normal voice transmission occupies at least a 4-kc bandwidth, it is apparent that the amount of traffic carried on in a particular spectrum is dictated, at the maximum, by the number of 4-kc channels available. For long distance communications which use the high frequency spectrum, we could provide, ideally and theoretically, approximately 8000 voice channels. However, due to practical considerations and military requirements, this number might be reduced to between 4000 and 6000 channels. If it were possible to provide four times as much voice traffic on the same number of channels, we would be essentially buying spectrum, a commodity rarely put up for sale; utilization of speech compression techniques would enable us to provide this additional capability. By means of multiplexing compressed speech, it is possible to place several voice channels in the spectrum of 3 kc, which is the nominal bandwidth requirement for normal speech as transmitted over existing radio links or wireline voice circuit facilities.

Another obvious military need in the communications field is that of providing proper security for information transfer; to this end also, speech compression can be utilized. To provide the best type of security, digital techniques are used. When normal speech is digitalized, the required bandwidth for transmission is on the order of 32 kc, or an approximate increase of 8:1. If speech is compressed first, and then digitalized, provision for transmission of encrypted digital speech over a normal 3-kc voice circuit can be made. An additional factor in providing speech in digital form is the increased reliability attendant to digital transmission schemes. Also, a certain degree of privacy can be achieved in either the analog or digital modes of transmission because, in order to receive intelligence of the transmitted speech, one must have the proper receiving equipment to decode and reconstruct the speech.

From the preceding it can be seen that speech compression will allow a few of the requirements set forth by the military in the communications field, primarily, bandwidth conservation and security, to be fulfilled. Techniques to satisfy these requirements have been available for some time; however, equipments compact enough and capable of producing a sufficiently good quality of reconstructed speech have not been available as acceptable end items. However, the Hughes Aircraft Company has developed a miniature Vocoder of sound quality which can be utilized for the above outlined military applications. This equipment will be described in detail later.

Compression Techniques

The word Vocoder is a contraction of the words "voice" and "coder," and the Vocoder system is an analysis-synthesis system of voice communications. Vocoder equipments were developed from an analysis of the mechanism of human speech production, and serve to reduce the bandwidth required to transmit speech. This is done by developing descriptive signals of significantly less redundancy than the original speech, while retaining voice intelligibility in the reconstructed speech. The Vocoder has evolved due to the results of intensive studies of the production and physical nature of speech conducted by Harvey Fletcher in the mid-1920's and by Homer Dudley in the 1930's. Their investigations resulted in a better understanding of the part played in the production of speech by the vocal chords, lips, tongue, teeth, nasal passages, and lungs. The studies indicate which parts of speech are redundant and which parts are significant in conveying intelligibility in order that a good replica of the original speech could be artificially constructed. The Bell Laboratories have been continuously involved in the advancement of the state-of-the-art of speech compression techniques and analysis, as have several universities.

The information in speech is carried largely in the varying shape of the power density spectrum and not in the sound pressure vs time plot seen on an oscilloscope. In the fixed-channel type of Vocoder, an analysis of the power spectrum is made by means of a series of bandpass filters. These filters are selected to divide the audio spectrum into n adjacent bands. An analog measure of the energy in each of these filter bands is obtained by envelope-detecting the output of each filter. Thus, n analog signals are developed which give a rough, but continuous, description of the frequency spectrum of incoming speech. These are the spectrum channels which convey the primary information contained in the speech input. The Vocoder continuously analyzes speech for two types of information, the spectrum channel information, as described above, and pitch channel information. Speech is composed of "voiced" sounds and "un-

voiced" sounds; the voiced sounds are larynx-type tones produced in emitting vowel sounds and are generated by the vibration of the vocal chords under pressure of air from the lungs. Voiced sounds are composed primarily of harmonics of the frequency at which the larynx vibrates. The fundamental frequency is approximately 70 to 250 cps for males, and as high as 350 cps for females. The "unvoiced" sounds are the breath sounds and are formed by the lips, teeth, tongue, and nasal passages in producing consonants. Unvoiced or consonant sounds have no definite harmonic frequency pattern, but are composed of frequencies randomly distributed throughout the spectrum and varying in amplitude in accordance with the sound being produced. Generally, voiced and unvoiced sounds are not present at the same time; however, sounds which do include both voiced and unvoiced sounds are treated as voiced sounds in the Vocoder with very little error. The pitch extractor circuit generates a control voltage for a relay at the receiving end connecting the output of a "buzz" generator, which serves to reinsert the voiced components of the speech, or a "hiss" generator, which reinserts the random frequencies contained in the unvoiced sounds. To synthesize the voice, a similar series of bandpass filters is used in conjunction with the outputs of the buzz and hiss generators and balanced modulators to produce the reconstructed and intelligent speech.

There are other types of speech compression techniques which will be described briefly as they have not yet achieved the high degree of equipment specification capabilities as the fixed-channel analysis-type Vocoders.

The systems that most nearly approach the performance of a fixed-channel analysis Vocoder are those of Formant Tracking Vocoders. A formant frequency is the frequency of a normal mode of vibration of the human vocal mechanism. During the utterance of certain sounds, notably the vowels, the normal modes of vibration of the vocal system are manifested in the acoustic output of the speaker as maximum in the spectra of these sounds. The formant frequencies then are the frequencies of the spectral maximum. The primary information analyzed by a Formant Tracking Vocoder is the frequencies and amplitudes of formants of the speech input. Such a system is capable of transmitting the vowel sounds of continuous speech fairly adequately. It is also capable of transmitting some consonant information. With this technique, compression ratios of 20:1 are possible but, unfortunately, the intelligibility as contrasted with fixed-channel analysis techniques is rather poor; articulation tests indicate scoring of about 65 percent or less.

There are also techniques utilizing hybrid compression systems. These are attempts to combine the advantages of the narrow bandwidth achieved with pure formant tracking techniques and high articulation achieved by systems with provisions for consonant synthesis, such as the fixed-channel Vocoder. One of these is the continuous-parameter type. This system makes use of the fact that speech possesses certain small numbers of parameters or characteristics which vary with time to produce the wide variety of speech sounds. These characteristics are measurable physically and describe the acoustic effect of the vocal tract on the source. They vary at a slow rate due to the physical inability of vocal mechanisms to adjust very rapidly. These low frequency signals may be transmitted in a standard manner and, when received, used to control a synthesizer incorporating sources similar to the physical sources which produced the original sound. The accuracy with which these parameters are extracted affects intelligibility as well as quality. This system exhibits greater intelligibility than a pure formant-tracking system, but the bandwidth is affected critically by the intelligibility desired.

Another hybrid technique that shows promise is a combination of fixed-channel analysis and formant-tracking techniques. In this scheme, perhaps six to eight fixed channels for analysis would be provided which cover the upper frequencies of the speech spectrum, plus a formant-tracking section to extract information of the first two formant frequencies and of the amplitude of the formants. No qualitative data is available on this technique. It could be expected to operate with good intelligibility within a bandwidth of approximately 250 cps.

Another hybrid system contains straight speech in the lower channels and Vocoder techniques in the upper channels. This offers intelligibility on the order of 80 percent by articulation scoring but with a wide bandwidth of upwards of 800 cycles. Straight band-chopping systems simply limit the speech bandwidth by removing the upper frequencies with filters. Since much of the consonant information is contained in the upper frequencies this system is very poor in intelligibility.

Description of Hughes Vocoder

In an effort to provide the military with equipments which encompass the requirements set forth, Hughes Aircraft Communications Division, has developed a Vocoder which has the necessary intelligibility and meets the environmental conditions for military equipments.

Two basic packages have been designed. One is an airborne Vocoder that operates in the simplex mode (see Figures 1 and 2). This unit consists of two suitcase-type aluminum boxes.



FIGURE 1. AIRBORNE VOCODER, COMPLETE.

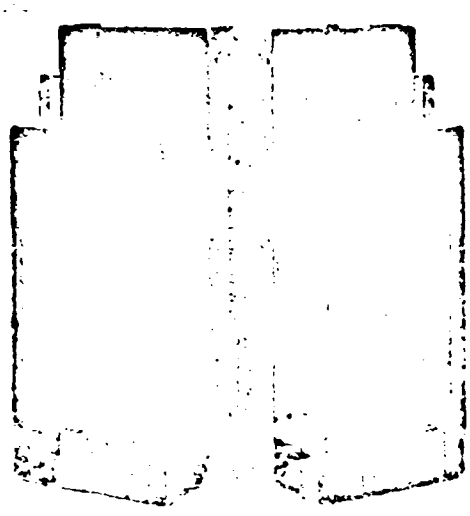


FIGURE 2. AIRBORNE VOCODER, BOX 1, MODULE SIDE.

Each box has over-all measurements of approximately 19-1/2 × 10 × 6-1/2 inches with irregularities as seen in the photograph. The approximate combined volume is 1.3 cubic feet and the total weight is less than 60 pounds. The power consumption is less than 100 watts ac. Also, a duplex model for point-to-point transmission has been built. This unit consists of three drawers, each 3-1/2 inches high, 19 inches wide, and 23 inches deep, fitted into a standard rack, (see Figure 3). The total weight, including power supplies, is about 200 pounds. The power consumption is 150 watts. These units provide circuits for digitizing speech for transmission as well as the decoding circuits for conversion of the digital information to synthetic speech.

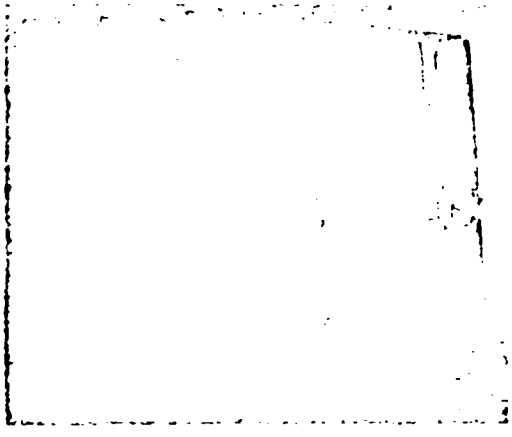


FIGURE 3. GROUND VOCODER, COMPLETE

All circuits are transistorized and are packaged in the form of "cordwood modules," (see Figures 4, 5, 6, and 7). The construction consists of two printed boards mounted parallel to each other in the manner of a sandwich with the components disposed between the boards in the manner of cordwood. The board wiring is made by the etching process either upon paper phenolic or epoxy glass base. Wiring appears on one side of the board only and all holes are dispersed on a grid of 2/10-inch centers. Wiring holes are countersunk on the interior of the boards to facilitate assembly. The assembly is given a protective coating to provide a dirt and moisture barrier for the printed circuitry. By presenting a low profile (1.75 inches high), and a relatively broad base, a high degree of module stability under the conditions of vibration and shock is achieved. In addition, this type of a subassembly is inherently less susceptible to damage resulting from field handling. The entire Vocoder has been transistorized with the exception of one subminiature thyratron used as a noise source. The ground duplex Vocoder



FIGURE 4. TYPICAL CORDWOOD MODULE.

uses 330 transistors and 1325 diodes, while the airborne unit uses 250 transistors and 915 diodes. All transistors used are silicon, and all individual circuits have been designed and tested for temperature environments of -55°C to +100°C.

Figures 8 and 9 are simplified block diagrams of the Vocoder divided into transmit and receive portions. On the transmit side the audio input is applied to the Voice-Operated Gain-Adjustment Device (VOGAD) in which the speech volume is adjusted automatically to a level approximately uniform for all speakers. The output of the VOGAD is applied to 12 spectrum channel filters which divide the audio input of 200 to 4100 cps into 12 frequency bands. As indicated in Figures 8 and 9, the bands increase in width from 160 cps for the lowest to 755 cps for the highest frequency





FIGURE 5. TYPICAL HALF SIZE CORDWOOD MODULE.

band. These 12 spectrum channels convey the articulation of the speech input. The outputs of each filter are fed to a spectrum analyzer which is composed of amplifiers and rectifiers. The output of the VOGAD is applied to the pitch extractor circuitry which produces the control signal which, when received at the receive terminal, synthesizes the speech through the proper application of either the buzz or hiss generators.

Once the analog information is made available in the form described above, facilities may be provided for either multiplexing several analog Vocoders on one transmission line or feeding these data to an analog-to-digital converter to produce digitized speech. The analog information on each analysis channel is defined by eight levels with three bits, whereas the information on this pitch channel is defined by 64 levels with six bits. This binary information is generated at the rate of 2000 bits/second. Each of the 13 channels is sampled every 21 ms.

The function of the receive analog section is to synthesize the speech from local sources of energy, or the buzz or hiss generators, described previously. To reiterate, the buzz generator provides the energy for the voiced or vowel sounds, and the hiss generator supplies the energy for the unvoiced or consonant sounds. The decision relay is controlled by the pitch control signal. The outputs of the generators (either buzz or hiss) are connected to the 12 spectrum synthesizer filters which are similar to the spectrum filters used at the transmit terminal. The outputs from the spectrum synthesizers are passed through bandpass filters to remove unwanted modulator distortion products, combined into one signal which is applied to the output amplifier, and thence to a speaker.



FIGURE 6. DIODE GATING STICK ASSEMBLY.

In the digital mode the digital data received are fed through digital-to-analog converter circuitry to produce the appropriate analog voltage defining each analysis channel. The analog output is then treated in the same manner as described previously to produce synthesized speech.

Extensive articulation testing of the Hughes Aircraft Company Vocoder has been done

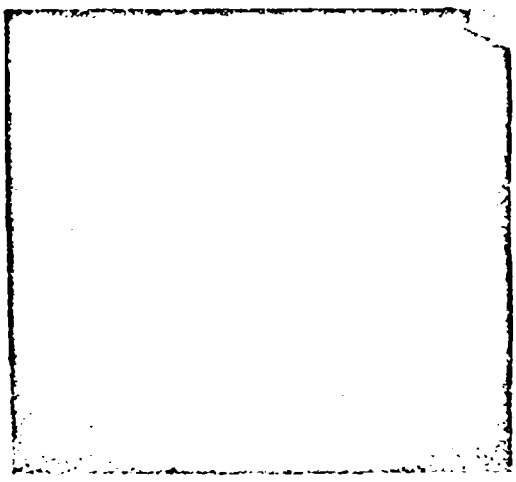


FIGURE 7. MODULE AND GATING STICK ASSEMBLY.

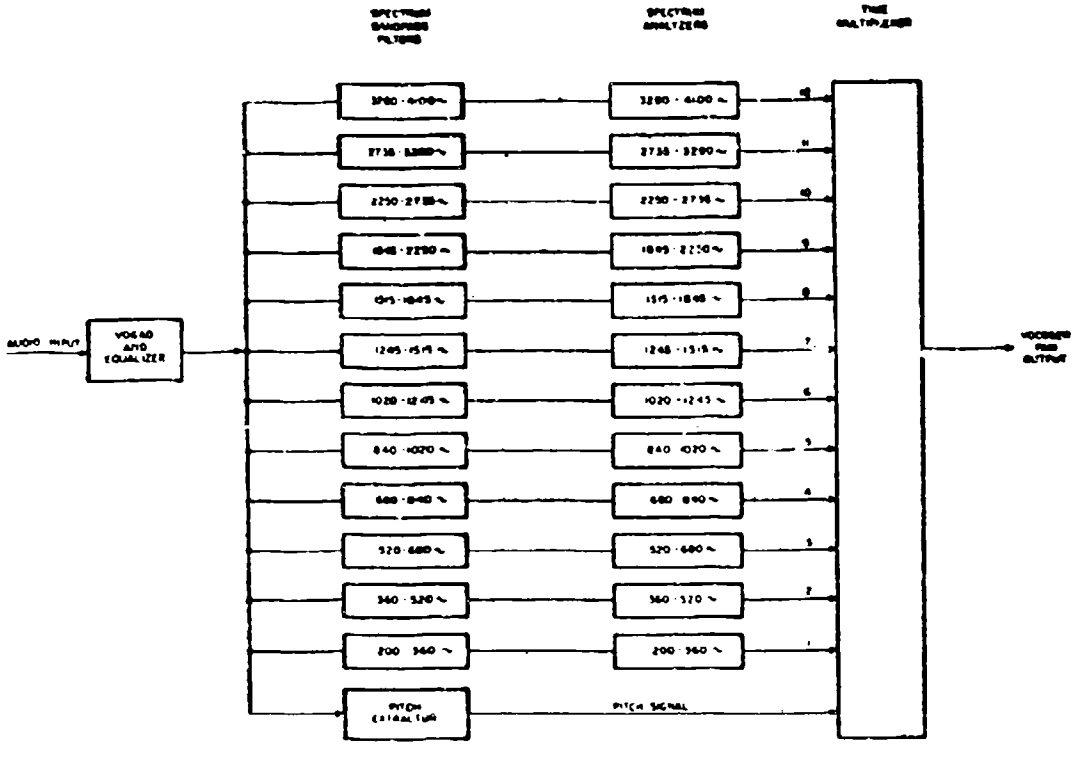


FIGURE 8. VOCODER TRANSMIT SECTION, BLOCK DIAGRAM

utilizing standard PB word lists for testing in both the analog and digital modes. Articulation scores of up to 90 percent have been achieved indicating a sentence intelligibility factor of almost 99 percent as shown in Figure 10. In the digital mode, controlled bit error rates were introduced into the bit stream of the digitized voice and tentative results show that a 4-percent error can be tolerated and still produce highly intelligible speech.

Utilization

Analog Mode

The basic analyzer portion of the Vocoder generates signals of extremely narrow bandwidth, 25 cps or less, on several wires. The Vocoder's bandwidth-compression capability may be utilized to the fullest extent by time-multiplexing samples of these signals and transmitting the resulting pulse amplitude modulated signal. This PAM signal derived from the Vocoder may have a bandwidth as narrow as 250 cps. In some applications, this 250-cps signal may be transmitted over a link as is, however, an attractive utilization is the multiplexing of several of these signals from different Vocoders into one composite signal. This composite signal may then be transmitted over a link presently designed for speech. In effect, the number of voice channels in existence would be multiplied without creating new links, but merely by adding equipment at the terminals. Figure 11 indicates the basic block diagram for this technique.

Using vestigial sideband techniques and frequency multiplexing, maximum use of bandwidth may be achieved. Including guard-bands, about 500 cps per voice channel is required, so that up to six voice channels may be transmitted over an audio link of 3-kc bandwidth (see Figure 12). A signal-to-noise ratio of about 25 db, peak signal to rms noise, is adequate for this composite signal. Essentially, the composite signal may be treated similarly to a normal speech signal with the exception of requiring a slightly better signal-to-noise ratio, and no compandors used in the link.

The ability to pass the composite signal over a single sideband link allows the possibility to frequency multiplex several of these signals with equipment identical to that which is used presently to multiplex normal speech signals.

Digital Mode

Probably a more important employment of Vocoders is in the conversion of a speech signal into a digital bit stream. Speech can, of course, be digitalized directly using pulse code modulation or delta modulation. However, to do so requires

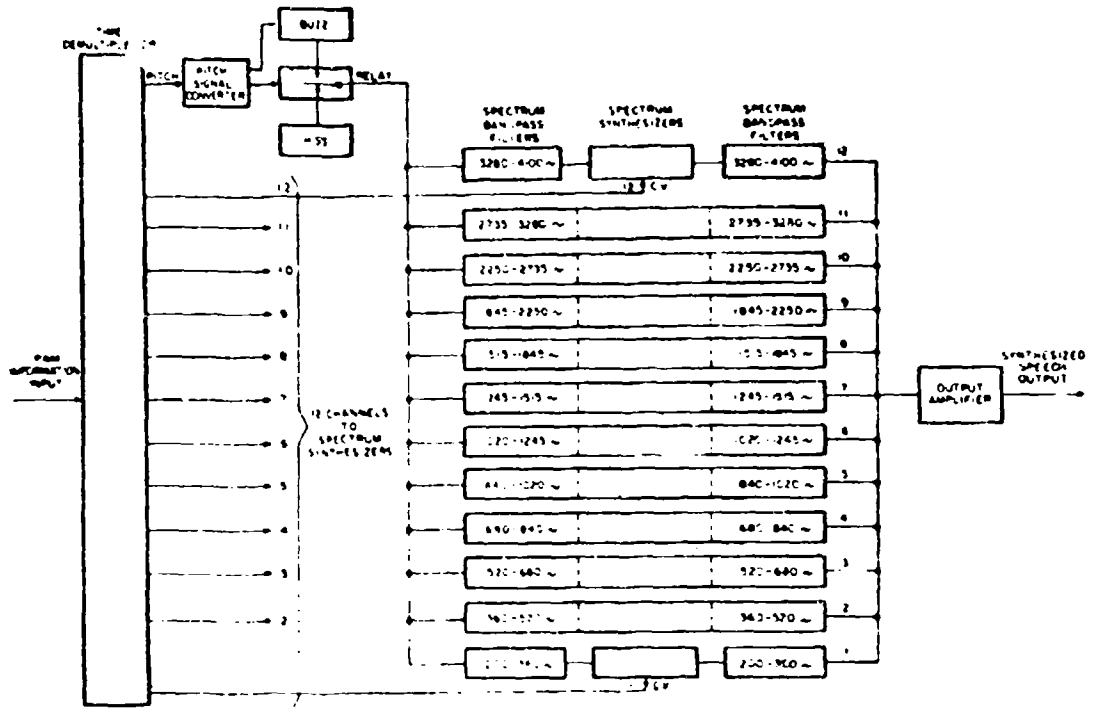


FIGURE 9 VOCODER, RECEIVE SECTION, BLOCK DIAGRAM

20,000 to 50,000 bits/second to achieve adequate speech quality. It is seldom feasible economically to transmit such a high bit rate for only one voice conversation.

Conversion of signals from the Vocoder analyzer portion into a digital bit stream requires only in the order of 2000 bits/second. The resultant Vocoder bit stream is not degraded readily by errors; digital error rates as great as 1 or 2 percent have a negligible effect upon speech intelligence. Such error rates in a normal teletype system would cause up to 10 percent of the characters to be erroneous, seriously affecting the readability of a message. Conversion of speech into a digital bit stream allows far more versatility than the original speech for signal transmission.

Various types of phase modulation or frequency modulation techniques may be used to transmit the bit stream in addition to amplitude modulation or vestigial sideband modulation. Sidestepping schemes may be utilized to eliminate multipath problems with high frequency transmission, multichannel devices may be used to overcome phase distortion over extremely poor audio links, allowing the signal to be transmitted with a significantly higher average to peak power ratio than that of ordinary speech, causing the transmitting signal to be far less subject to interference.

The ability to encrypt the digital bit stream provides as high a level of secure transmission of speech signal as may be desired. Voice scramblers provide privacy; however, a Vocoder operating with a cryptographic device allows the greatest security obtainable.

Serial multiplexing of several Vocoder bit streams can be achieved simply at a very low cost; whereas, to multiplex say 24 voice channels into a 100-kc bandwidth now requires 24 single sideband modulators and numerous filters, 24 digital Vocoders, each running at 2 kb/second, can be multiplexed into a single 48-kb bit stream using only six flip-flops. This 48-kb bit stream may be easily modulated, with a single modulator, into a signal with a bandwidth of less than 100 kc. In fact, using vestigial sideband techniques, 32 Vocoder outputs can be multiplexed into the 100-kc bandwidth with no advance in the state-of-the-art. Thus, as many as 32 fully encrypted speech channels may be transmitted with a substantially higher average to peak power ratio where only 24 ordinary speech channels are now in operation, with much simpler multiplexing equipment.

Conclusions

Up to the present time Vocoders have not seen much use in military or other applications. Some Vocoders were used in World War II and, since, by high level people in government for security aspects. Heretofore, Vocoders have been plagued with several shortcomings, large physical size, excessive weight, high power, poor intelligi-

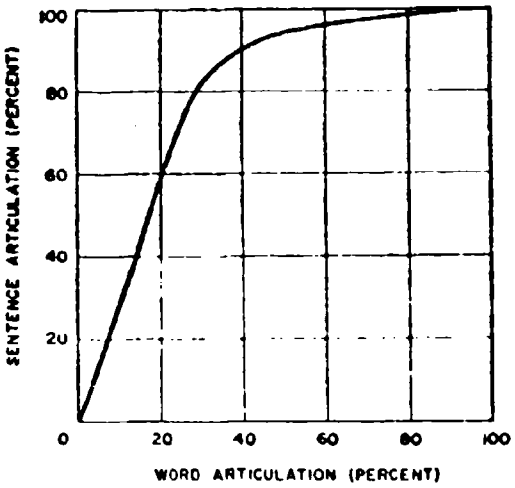


FIGURE 10. WORD-TO-SENTENCE ARTICULATION TRANSFER.

bility, sensitivity to different speakers, and poor speaker recognition capability. These shortcomings are being eliminated. Hughes Aircraft Company has packaged a transistorized digital Vocoder with word articulation of 85 to 90 percent for speech digitized at 2000 bits/second. This results in near-perfect sentence intelligibility. With recent developments, sensitivity to different speakers has been greatly decreased; it is expected that speaker recognition will soon be as great as may be desired.

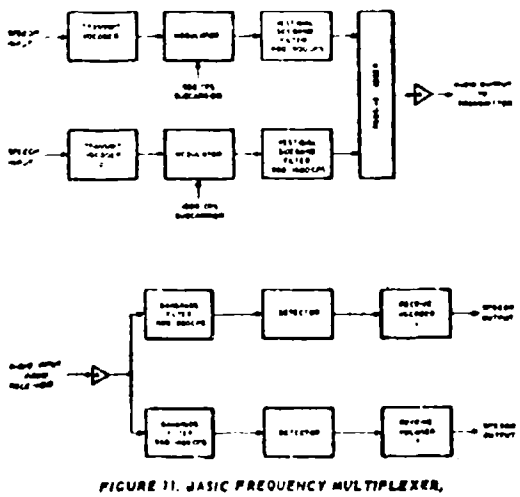


FIGURE 11. BASIC FREQUENCY MULTIPLEXER, SIMPLIFIED BLOCK DIAGRAM.

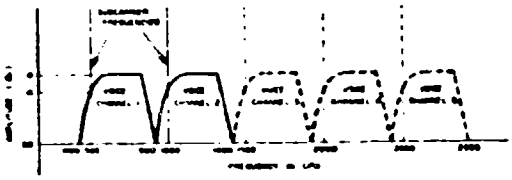


FIGURE 12. VOICE CHANNEL FREQUENCY ASSIGNMENTS, BANDPASS CHARACTERISTICS.

Bibliography

Benoit, Richard C. Jr., "A USAF Multiplex Equipment Development," September 1959

Dudley, Homer, "The Carrier Nature of Speech," Bell System Technical Journal, Vol. 19, 1940.

Dudley, Homer, R. R. Riesz, and S. S. A. Watkins, "A Synthetic Speaker," Journal of the Franklin Institute, Vol. 227, 1939.

Dudley, Homer, "The Vocoder," Bell Laboratory Record, Vol. 18, 1939.

Dudley, Homer, "Remaking Speech," Journal of the Acoustical Society of America, Vol. 11, 1939.

Flanagan, James L. and Arthur S. House, "Development and Testing of a Formant-Coding Speech Compression System," Journal of the Acoustical Society of America, Vol. 28, No. 6, 1956.

Flanagan, James L., "Automatic Extraction of Formant Frequencies From Continuous Speech," Journal of the Acoustical Society of America, Vol. 28, No. 6, 1956.

Fletcher, Harvey, "Speech and Hearing in Communications," D. Van Nostrand Co., 1955.

Howard, Calvin R., "Speech Analysis-Synthesis Scheme Using Continuous Parameters, "Journal of the Acoustical Society of America, Vol. 28, No. 6, 1956.

Hughes Aircraft Company, Report R-114, "Twelve-Channel Transistorized Vocoder."

Sharf, Donald J., "Intelligibility of Reiterated Speech," Journal of the Acoustical Society of America, Vol. 31, No. 4, 1959.

Weinstock, M., "The Army and Speech Processing," Seminar Proceedings on Speech Compression and Processing, Vol. 2, September 1959.

TOPOLOGY ENGINEERING OF COMMUNICATIONS NETWORKS

Kurt Ekrath and Charles C. Comstock
U. S. Army Signal Research
and Development Laboratory
Fort Monmouth, New Jersey

ABSTRACT

This paper discusses the topological parameters which express the correspondence relations between communication network layout organizational structure, the communicators, the system operating principles and the systems supervision and housekeeping rules, and criteria regarding flexibility, versatility, expandibility, and vulnerability of communications networks. Practical illustrative examples of topology engineering in connection with the synthesis and design of a Universal Integrated Communication System are given to show that the design of such a system can conform to a synthesis philosophy which yields a complete understanding and control of the relations and dependencies between system elements rather than to improvise concepts and doctrines, which when forced to changed, render the system obsolete.

1. INTRODUCTION

A communication system is always an organic part of a related larger human-machine system and it is established for the support of and service to this larger system. Because of its human population the larger system is in a transitory state which is evidenced by the rapid obsolescence of certain human missions and or machine functions, and their underlying social doctrines and operational concepts. Among the organic parts of the larger system, the communications system is a controlling factor with the unique capability of rendering the whole larger system inoperative. While the missions and functions of communications per se is never endangered by obsolescense as long as the human population exists, communications systems have become obsolete in the past. This rate of obsolescence of communications systems is most pronounced in all cases where the communication system was specifically tailored to specific requirements of a small class within the human machine population of a larger system. In these cases the models of the communication systems were related to selected realities of the larger system as they existed at a time previous to the conception of the model. Thus one is justified in calling the art of communication system development, "Retarded Development." The question arises if Communications system development must forever be retarded. That it need not be can be demonstrated by uncovering the canonical elements and forms of a universal communication system model. (Canonical means the invariants of the fundamental relationships and dependencies between system elements relative to changing frames of their deployment and applications.) This implies that one must not view a communication system from a specific outside point within the larger system, but one must view a communication system from within against the larger system as background. In other words, by formulating the intrinsic characteristics of communication systems a universal system model is attained which can be projected into any deployment frame and applications scale.

2. TECHNICAL BACKGROUND

This paragraph is devoted to an explanation of the terminology which is used to describe the purpose, functions, and objectives of the systems model as well as concepts, methods, and techniques used for implementing the model.

The classification of Models.

A system S' is called a model of the system S, if given certain facts about system S, other facts of S can be inferred from the model S'. A system S' is also a model of S, if from postulates or assumptions associated with S' facts of S can be inferred. In the first case elements of S' correspond to observable elements of S and the model S' is thus representative for S. In the second case the elements of S' need not correspond with observable elements of S and the model S' is thus non-representative for S.

If a Model is representative and its elements are organized in exactly the same way as the system S that it represents, then the model is called iconic. (e.g. S' is a small scale replica of S.)

A model S' of S is an analogue model if it utilizes physical phenomena, devices, and pictorial drawings for the modelling of S. As such it is usually aimed at being a representative model of S.

A model S' of S is a symbolic model if it utilizes abstract elements and operations for the deduction of facts about S. A symbolic model is never a representative one. A symbolic model may use deterministic or probabilistic concepts and approaches for the modelling of the system.

The selection of or conception of a communications system is intimately linked with the familiar criteria or measuring communications system effectiveness. These are reliability, accuracy, capacity, speed, flexibility, mobility, invulnerability, security, connectivity, efficiency, and cost.

The selection of or conception of a model on the other hand is governed by the familiar criteria of faithfulness, complexity, generality and availability, physical size, fastness and cost. In practice most models of actual systems use more or less all of these features explicitly or implicitly. Recognition of the va-

lidity range of a model depends to a great extent on the capability of discriminating between these features of the model. While the meaning of these terms are known and generally accepted, faithfulness of a model constructed to imitate, describe or understand natural phenomena and naturally observed effects is clearly determined by the degree of agreement between model derived effects and naturally observed effects. Faithfulness of a communications system model implies the obsolescence of the subsequently implemented communications system, since this faithfulness refers as pointed out in the introduction to a passed state of the larger system. Thus, faithfulness to the requirements of restricted classes of communication system subscribers, e.g. military subscribers, which have been conditioned to act in certain ways by furnishing them with systems faithful to their prior requirement will result in faster and faster rates of the obsolescense cycles. Only a universal system model where faithfulness in the above sense is traded for faithfulness to emphasized fundamental properties has any chance of leading to a system which is not plagued by obsolescense of doctrines and missions of its users.

The fundamental property of a system is the connectivity between its elements. The symbolic models which concern themselves with this connectivity are the topological models. Topological models represent the rubber-sheet geometry of the model system; and it is this rubber-sheet geometry, i.e. the topological parameter, of the system which have a pronounced influence on the design, operation, and performance of the system as will be shown in the following discussion.

3. DISCUSSION OF THE TOPOLOGICAL MODEL

A. THE CANONICAL ELEMENTS OF A COMMUNICATION SYSTEM.

In any communication system one can distinguish between the following subsystems: Subscriber subsystems, Central Office subsystems, and Transmission subsystems. However, the boundary between the subsystems need not be clearly delineated and may be subject to interpretation on the basis of techniques, methods, and devices used in the various subsystems. (e.g. TD multiplexing and TD switching which are similar in basic working principle can be lumped together or inter-dispersed to such an extent that their associated functions are physically identifiable with the central office and the transmission subsystem, respectively.

B. THE SYMBOLIC REPRESENTATION OF THE CANONICAL ELEMENTS.

The set of building blocks, i.e. elements which will be introduced consist of two configurations: One is the elementary graph of first degree shown below in Figure 1, the other is the elementary graph of second degree shown below in Figure 2.

Figure 1 Figure 2



These two building blocks play a role in communication network configurations which are similar to the units of information. Like the syntax of information is conceived of as an aggregate of bits, nits, etc; a communication network configuration can be conceived of as an aggregate of elementary graphs of first and of second degree. Lines symbolize the transmission subsystems, nodes stand for subscriber and/or central office subsystems sometimes called stations and exchanges respectively. Their positionin; relative to other nodes in a graph, i.e. their rank, determines their standing as subscribers or central office subsystems. As the symbolic meaning of Figure 1 and Figure 2 regarding communication subsystems is self evident the question arises in which way these elements enter into a communications network in terms of functional, operational, and performance criteria.

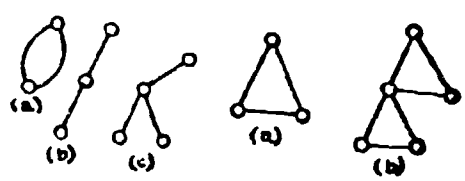
Communications systems are formed in practice by joining together transmission terminals into aggregates which are called subscriber subsystems and central office subsystems. This amounts symbolically to adding together elementary graphs in a way such that in the resultant graph a pair of elementary graphs may share the same nodes (coincidence of nodes) but do not have any of their branches in common (no extinction of branch lines when joining subgraphs.) For example two elementary graphs of first degree can be added together as indicated in Figure 3 or two elementary graphs of second degree are added as shown in Figure 4.

In any network or graph constructed in such a way the following relations are established between the number of branches b and the number of nodes N

$$2b = n_1 + 2n_2 + 3n_3 + 4n_4 + 5n_5 + \cdots = \sum_{v=1}^{s} v \cdot n_v \quad \text{where} \quad \sum_{v=1}^{s} n_v = N \qquad (1)$$

where n_v is the number of nodes of degree (i.e. rank) v. (branches emanate from each of the n_v nodes.)

Figure 3 Figure 4

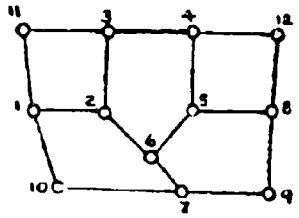


For example $2b = 2 \times 4 + 3 \times 8 = 32 \quad b = 16$

as illustrated below in Figure 5

Figure 5

Nodes Nr. 1-9 are of degree 3;
Nodes Nr. 9-12 are of degree 2



One may note that in any network (graph) the sum of nodes of odd degree must be an even number since 2b is even and one gets by deduction of an even

$$2s = 2n_2 + 2n_3 + 4n_4 + 4n_5 + 6n_6 + 6n_7 + \cdots$$

from $2b = n_1 + 2n_2 + 3n_3 + 4n_4 + 5n_5 + 6n_6 + 7n_7 + \cdots$

$$2p = 2b - 2s = n_1 + n_3 + n_5 + n_7 \quad (2)$$

which is an even number equal to the sum of all odd nodes.

While on the other hand a graph (network) with N nodes of mixed rank must contain $N - 2p = q$ nodes of even rank.

C. SIGNIFICANT FUNDAMENTAL STRUCTURES

Considering again the elementary graph of first degree as a basic building block, the number of branches B required to connect a set of N nodes such that only one unique path exists between any two nodes of the resultant graph follows with

$$\beta = N - 1 \quad (3)$$

Such a graph is then a fundamental structure, also called a tree of a network, as the network itself can be built up by further superposition of elementary graphs of first degree or second degree such that the number of nodes N is kept constant. (Coincidence of added nodes with original ones i.e. addition of branches) which produces a $b > \beta$. If the index of connectivity μ of a net is by definition $\mu = b - \beta$ then $\mu = 0$ indicates that the network is a tree where loss of a branch would result in the destruction of the connectivity within the set of nodes.

$\mu > 0$ indicates that up to μ branches can be removed from the network without destroying the connectivity subject to the constraint that the rank of every node remains larger or equal to one. Thus minimum connectivity is expressed by $\mu = 0$ with $b_{min} = \beta = N - 1$ (4)

The other extreme would be a maximum in connectivity which is clearly associated with a situation where every node is directly connected with every other node in the net by at least one branch. For one branch each the rank of every node is

$$\nu = \nu_{max} = N - 1 \quad \text{and} \quad n_\nu = n_{\nu_{max}} = N$$

giving $2b_{max} = \nu_{max} n_{max} = N(N-1)$ (5)

calling $b_{max} = \gamma$ the maximum number of single branches (excluding paralleling of branches) in a net of N nodes one gets

$$\gamma = \frac{N(N-1)}{2} = \frac{N!}{[(N-2)!][2!]} \quad (6)$$

equivalent to the combinations of taking N nodes 2 at a time.

And the maximum index of connectivity follows with

$$\mu_{max} = \gamma - \beta = \frac{(N-1)(N-2)}{2} = \frac{[(N-1)!]}{[(N-1-2)!][2!]} \quad (7)$$

(If every node is connected directly to every other node by a single branch)

The index of connectivity μ of the network shown in Figure 5 is

$$b - \beta = 16 - 11 = 5$$

also clearly associated with the 5 'Circles'
(1 - 2 - 6 - 7 - 10 - 1)
(1 - 2 - 3 - 11 - 1) (2 - 3 - 4 - 5 - 6 - 2)
(5 - 4 - 12 - 8 - 5) (7 - 6 - 5 - 8 - 9 - 7)
in the graph of Figure 5 which have in common with one another the branches (1 - 2) (2 - 3) (2 - 6) (5 - 5) (6 - 7) (5 - 4) (5 - 8).

The permissible numbers of nodes of different degrees n_ν in an ensemble of N nodes which can be connected to form a fundamental structure or tree follows from formula (1) (2) and (3) by setting

$$2\beta = 2(N-1) = n_1 + 2n_2 + 3n_3 + 4n_4 + 5n_5 + 6n_6 + \cdots$$

$$N = n_1 + n_2 + n_3 + n_4 + n_5 + n_6 + \cdots$$

$$1 \leq \nu \leq N - 1$$

$$N \geq 2p = n_1 + n_3 + n_5 + n_7 \quad (8)$$

The following table yields the n_ν which for given N satisfy the set of formulas (8) above, up to $\nu_{max}=10$ i.e. trees containing N=11 nodes. The underlying possible configurations are shown for the case N=11 is figures 6 (1) through (9). One may note that the configurations shown involving $(n_1,n_2\cdots n_9)$ are not necessarily unique e.g. different tree configurations of identical sets of n_ν called hence forth different members of the tree family $(n_1,n_2\cdots n_9)$ are shown in Figure 7 and 8. In the following tables σ is the number of significantly different members in every family of trees (Significantly different means that members which are symmetric or transitive versions of other members are not counted as different; this means that no distinction is made between the branches of the graph as will be done in Section 4.)

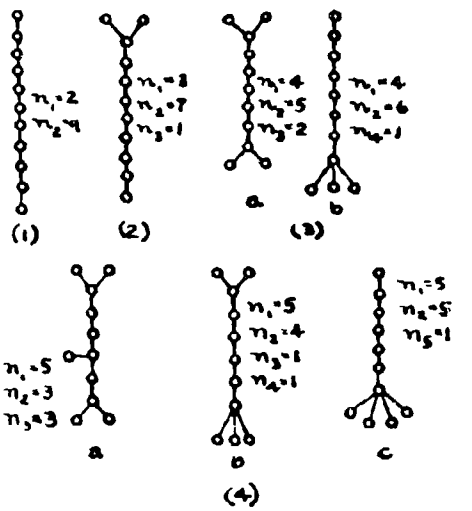
γ is the number of significant tree configurations which can be formed with N nodes and ϕ the number of families of trees for given N.

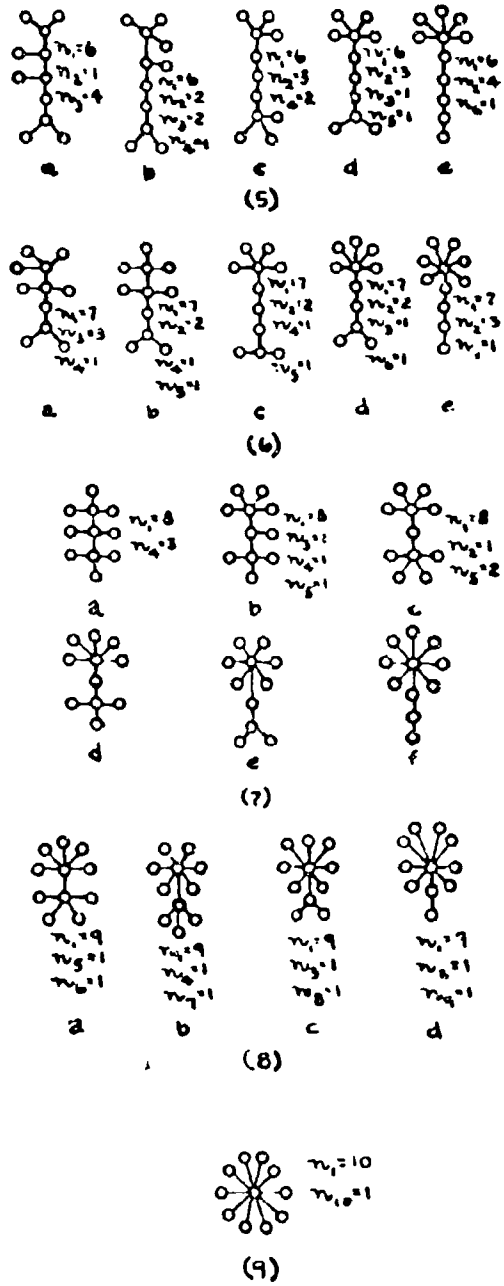
n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}	β	σ	γ	ϕ	N
2	7	0	0	0	0	0	0	0	0	1				
3	5	1	0	0	0	0	0	0	0	1				
4	3	2	0	0	0	0	0	0	0	1				
4	4	0	1	0	0	0	0	0	0	1				
4	1	3	0	0	0	0	0	0	0	1				
4	2	1	1	0	0	0	0	0	0	1				
5	3	0	0	1	0	0	0	0	0	1				
5	1	0	2	0	0	0	0	0	0	1	2	25	14	9
6	0	2	1	0	0	0	0	0	0	1				
6	1	1	0	1	0	0	0	0	0	1				
6	2	0	0	0	1	0	0	0	0	1				
7	0	0	1	1	0	0	0	0	0	1				
7	0	1	0	0	1	0	0	0	0	1				
7	1	0	0	0	0	1	0	0	0	1				
8	0	0	0	0	0	0	1	0	0	1				

n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}					
2	9	0	0	0	0	0	0	0	0	1				
3	6	1	0	0	0	0	0	0	0	1				
4	4	2	0	0	0	0	0	0	0	1	3	47	23	11
4	5	0	1	0	0	0	0	0	0	1				
5	3	1	1	0	0	0	0	0	0	1	2	32	16	8
5	4	0	0	1	0	0	0	0	0	1				
5	2	3	0	0	0	0	0	0	0	1				
6	3	0	0	0	1	0	0	0	0	1				
6	2	1	0	1	0	0	0	0	0	1				
6	2	0	2	0	0	0	0	0	0	1	3	57	21	10
6	1	2	1	0	0	0	0	0	0	1				
6	0	4	0	0	0	0	0	0	0	1				
7	1	0	1	1	0	0	0	0	0	1				
7	0	2	0	1	0	0	0	0	0	1				
7	0	1	2	0	0	0	0	0	0	1				
7	2	0	0	0	0	1	0	0	0	1				
5	1	0	0	0	0	0	1	0	0	1				
8	0	0	0	2	0	0	0	0	0	1				
8	0	0	1	0	1	0	0	0	0	1				
8	0	1	0	0	0	1	0	0	0	1				
9	0	0	0	1	0	0	0	1	0	1				

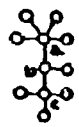
n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}	β	σ	γ	ϕ	N
2	9	0	0	0	0	0	0	0	0	1				
3	7	1	0	0	0	0	0	0	0	1				
4	5	2	0	0	0	0	0	0	0	1	3	47	23	11
4	6	0	1	0	0	0	0	0	0	1				
5	4	1	1	0	0	0	0	0	0	1	2	32	16	8
5	5	0	0	1	0	0	0	0	0	1				
5	3	3	0	0	0	0	0	0	0	1				
6	4	0	0	0	1	0	0	0	0	1				
6	3	1	0	1	0	0	0	0	0	1				
6	3	0	2	0	0	0	0	0	0	1	3	57	21	10
6	2	2	1	0	0	0	0	0	0	1				
6	1	4	0	0	0	0	0	0	0	1				
7	2	0	1	1	0	0	0	0	0	1				
7	1	2	0	1	0	0	0	0	0	1				
7	1	1	2	0	0	0	0	0	0	1				
7	3	0	0	0	0	1	0	0	0	1				
5	2	0	0	0	0	0	1	0	0	1				
8	1	0	0	2	0	0	0	0	0	1				
8	1	0	1	0	1	0	0	0	0	1				
8	1	1	0	0	0	1	0	0	0	1				
9	1	0	0	1	0	0	0	1	0	1				
10	0	0	0	0	0	0	0	0	1	1				

Figure 6 N=11





If one picks for example the graph 7b of Fig.6 shown below as one member of the family of graphs (n_1=8, n_3=1, n_4=1, n_5=1) and calls the inner nodes a, b, c, then we can associate the ranks ψ=3 ψ=4 ψ=5 with a, b, c, respectively yielding the following permutations of three elements and the associated 6 trees of Fig.7



a	5	5	4	5	4	3
b	5	4	5	4	5	5
c	4	5	5	5	5	4

ranks

Fig. 6, 7b

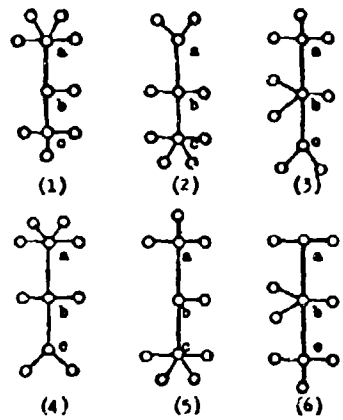


Fig.7

N = 11 (n_1= 8, n_3=1, n_4= 1, n_5= 1)

If one excludes symmetrical versions from being counted as significantly different graphs one obtains only three significantly different tree configurations because the trees numbered (1↓↑5) (2↓↑4) (3↓↑6) are as illustrated, upside down versions of one another. Similarly the graph (a) of fig.6 for instance, is one member of a family of trees (n_1=4, n_2=5, n_3=2) which comprises

$$P_1(5,2) + P_2(2,1) = \frac{7!}{5!\,2!} + \frac{3!}{2!\,1!} = 21 + 3 = 24$$

members as shown below in Fig. 8 (The upside down versions are indicated in the second row by ↓)

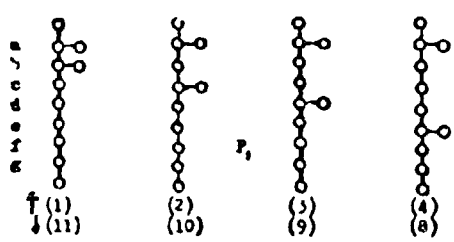
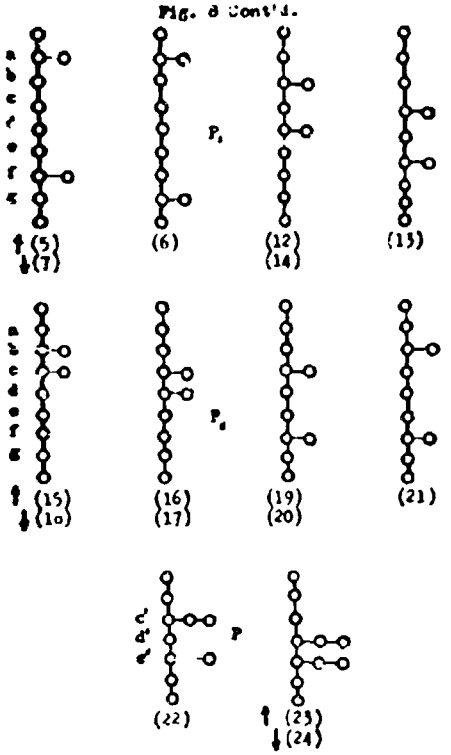


Fig.8

Fig. 6 Cont'd.



There are three symmetric cases among the P_1, namely (5) (13) and (21) of Fig.6 and one symmetric case among the P_2, namely (22). Thus the family (n_1=4, n_2=5, n_3=2) has 14 significantly different tree configurations as members. The members of the family (n_1=4, n_2=6, n_3=1) of Fig.6 are readily found from the following considerations: There are two symmetric cases in this family shown below in Fig. 9a and 9b.

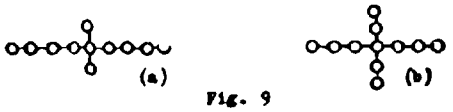
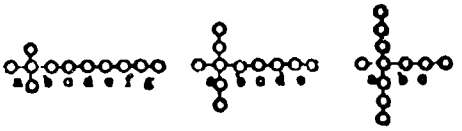


Fig. 9

From fig. 9a one obtains by shifting the node of rank four to the left (or to the right) three more significantly different trees (excluding left-right images). From fig. 9b one obtains by shifting the node of rank four to the left (or to the right) two more, and by shifting up (or down) one more significantly different tree. Thus one has $(\underline{1} + 5) + (\underline{1} + 2 + 1) = 8$ significantly different members in the family (n_1=4, n_2=6, n_3= 1). The underlined are those shown in Fig. 9a, 9b. Following the permutations rules one has

Fig. 9a Fig. 9b Fig. 9c



$P_1 = \frac{7!}{6!1!} = 7$ $P_2 = \frac{5!}{4!1!}$ $P_3 = \frac{3!}{2!1!}$

$P_1 + P_2 + P_3 = 7+5+3 = 15$ combinations where
among the $P_1 = 7$ are 2x3 mirror images
among the $P_2 = 5$ are 2x2 mirror images
among the $P_3 = 3$ are 2x1 mirror images
and among the combined $P_2 + P_1 = 8$ are 2x1 mirror images so that one gets 15 - 7 = 8 significantly different members.

Further sample illustrations for N = 6, N = 7, N = 8 are shown in Fig. 10a through 10e.

N =6 Fig. 10a

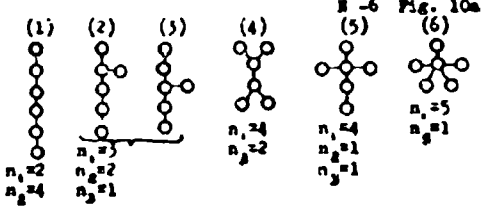


Fig. 10b N =7

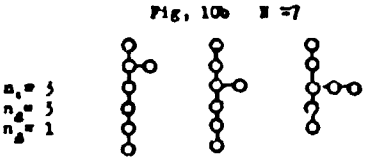


Fig. 10c N =8

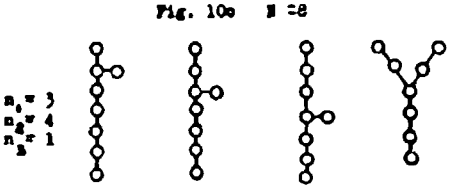


Fig. 10d N = 8

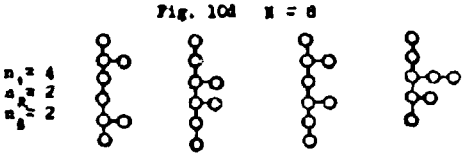
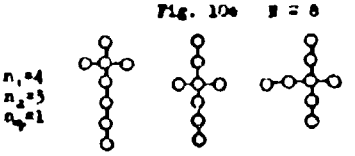


Fig. 10e N = 8



It is obvious that the inspection method of finding the significantly different members of a tree family for given n 's is rather tedious and becomes unmanageable for N beyond 11. However for all practical applications N - 11 (11 switching centers in a communications net) can be considered to be sufficiently large.

D. DISTINCT FUNDAMENTAL STRUCTURES.

Fundamental structures or trees whose elementary building blocks (elementary graphs of first degree) are labelled distinctly different and are called hereafter distinct. The question arises of how to find all possible distinctly different structures of a network with N nodes. The method of finding all distinct trees is demonstrated below in fig. 11a for a complete network with N = 4 nodes, each of rank 3

$$N = n_3 = 4 \text{ and hence } b_{max.} = \frac{1}{2} 4 \times 3 = 6 \text{ branches}$$

Fig. 11a



One selects one possible fundamental structure e.g. the original tree and derives all the others by successively eliminating branches 4, 5, 6 without losing connectivity, i.e. without producing a closed path or circle at the expense of cutting off a node. Thus, one obtains the following combinations.

Branch	4	5	6	Branches	4-5	4-6	5-6
with	1-3	2-1	1-2	with	1	1	1
with	2-3	3-1	2-3	with	2	2	2
				with	3	3	3

yielding the following 16 distinct trees in the complete net with four nodes.

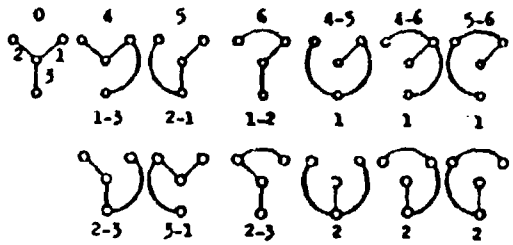
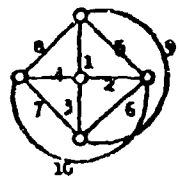


Fig. 11b

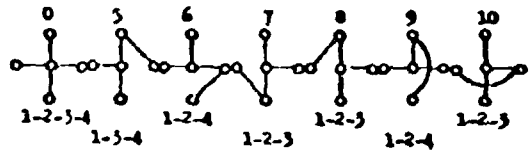


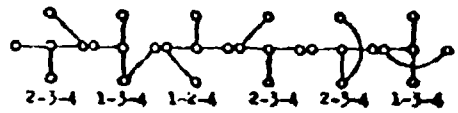
Analogously in a complete net with N = 5 nodes shown below in fig. 12a one may select the tree 1, 2, 3, 4 as the original starting point and obtain by successive elimination and substitution 125 distinct trees which are contained in the network.

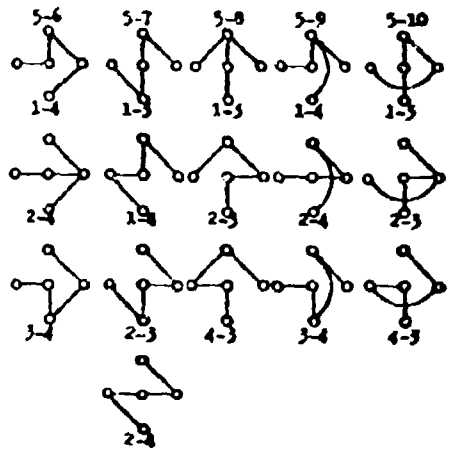
Fig. 12a

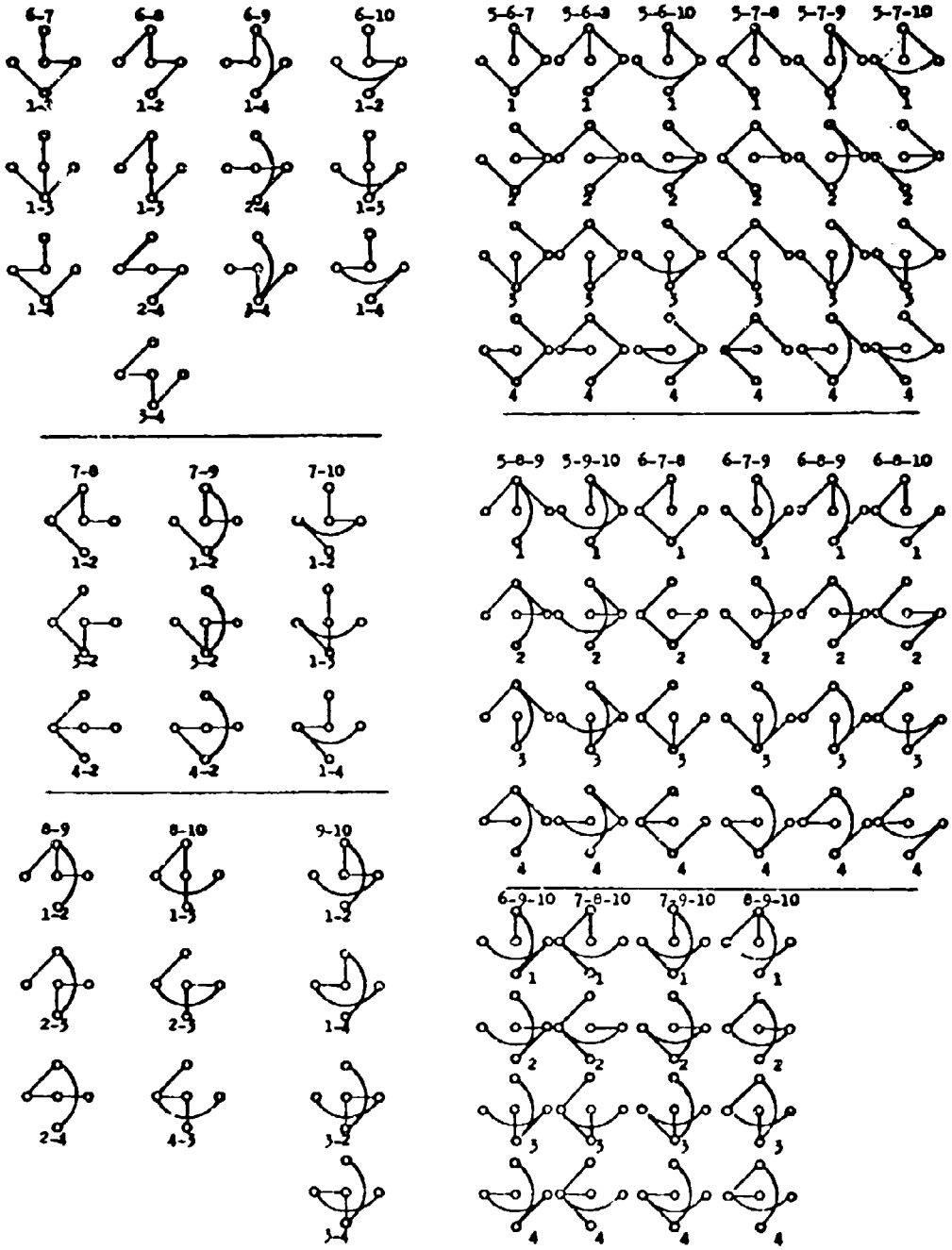


The possible distinct trees in terms of the combinations of four branches out of ten are shown in the following fig. 12b. One may note that combinations containing (5,6,9) (6,7,10) (5,8,10) (7,8,9) in connection with 1,2,3,4 respectively, leading to circles and loss of connectivity are not permitted.








6-7
6-8
6-9
6-10
5-6-7
5-6-8
5-6-10
5-7-8
5-7-9
5-7-10
7-8
7-9
7-10
5-8-9
5-9-10
6-7-8
6-7-9
6-8-9
6-8-10
8-9
8-10
9-10
6-9-10
7-8-10
7-9-10
8-9-10

The following regular network with N = 6 nodes with rank 4 i.e. N = n = 6 and b = 12 shown in fig. 13a also referred to as a "polar umbrella single belt network" contains 301 distinct tree configurations in terms of the permitted combinations of five branches out of 12.

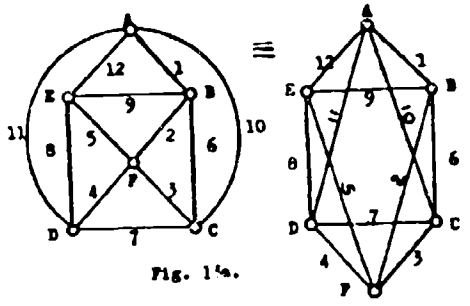


Fig. 13a.

For the derivation of the set of distinct trees the tree 1,2,3,4,5 has been selected as the starting point, as indicated in fig. 13b.

Among the non-permitted combinations which lead to circles are directly ascertained those combinations which contain (7-8-11-12)(7-8-10-12) (8-9-10-11-12)(8-9-11-12)(8-10-11-12)(7-9-10-11) (7-9-10-11)(6-9-10-12)(6-8-11-12)(6-7-10-11) (6-7-8-9)(9-11-12)(7-10-11) in connection with any one or any five branches out of the original set 1,2,3,4,5 respectively.

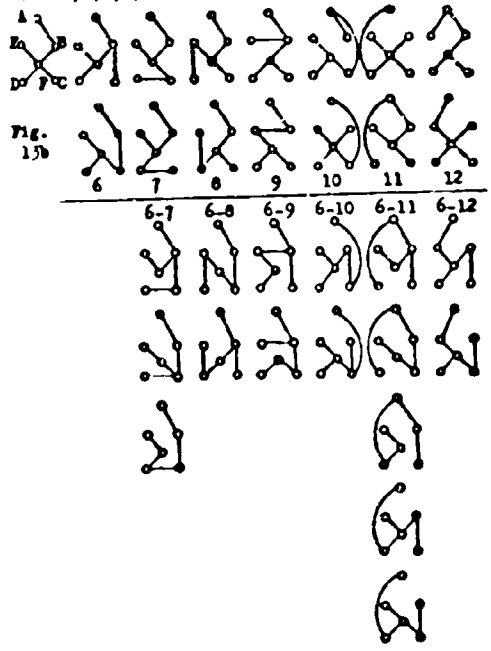
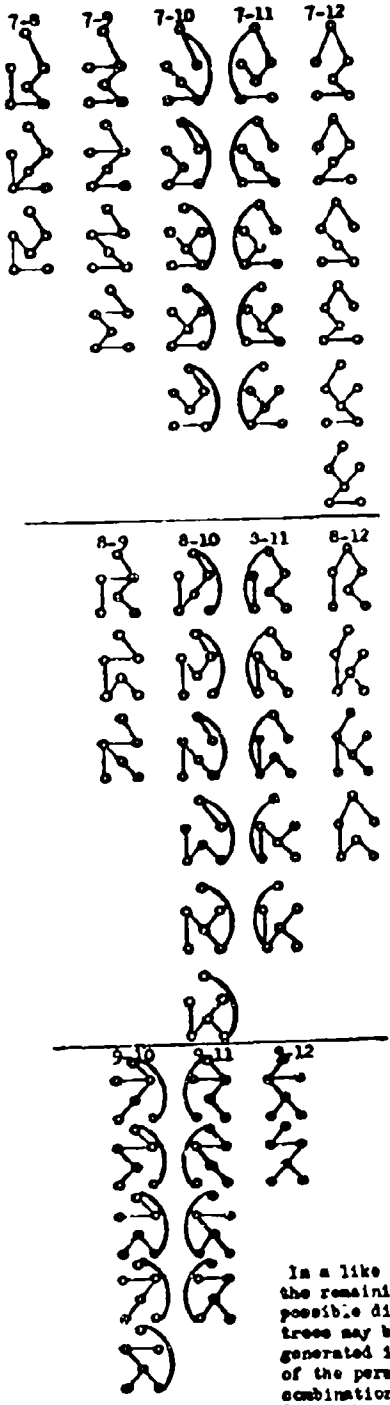


Fig. 13b



In a like manner the remaining possible distinct trees may be generated in terms of the permitted combinations of 5 branches out of 12.

E. REGULAR NETWORKS.

A regular network is by definition a network whose nodes are of equal rank or degree. Thus the complete network discussed before is a specialization of a regular network for [illegible]. A regular network of N nodes of rank ν contains b branches in accordance with formula (1) $2b = \nu N_\nu$ which requires that for N_ν being odd, ν to be an even number and vice versa for ν being odd N_ν to be even. Synthesis of regular networks in terms of a certain amount and types of canonical elements (regular graphs of first or second degree) is demonstrated by the following regular networks of figs. 14a, b, c, d with even N, and figs. 15a, b, c with odd N. In figs. 14a, b, c, d a distinction is made between each of the branches emanating from a node by coding. The objective of the synthesis is the matching of the coded branches in the formation of the network.



$\nu = 3$
$N = 4$

Fig. 14a

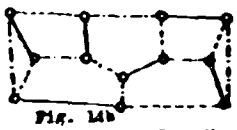


Fig. 14b

$\nu = 3$ $N = 12$

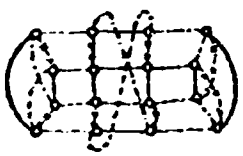


Fig. 14c

$\nu = 4$ $N = 16$

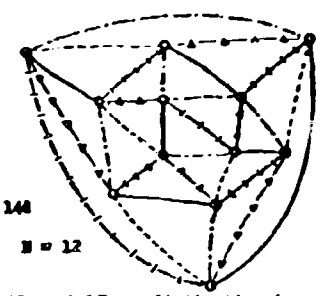


Fig. 14d

$\nu = 5$ $N = 12$

In fig. 15a and 15b a distinction is made between pairs of ν = even branches emanating from every one of the N_ν = odd nodes.

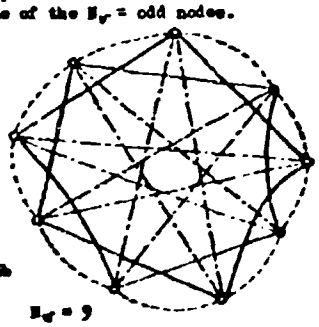


Fig. 15b

$\nu = 6$ $N_\nu = 9$

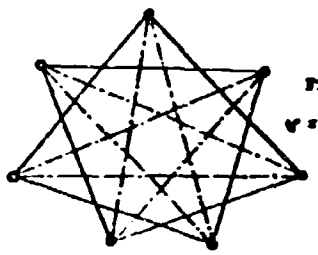


Fig. 15a

$\nu = 4$ $N = 7$

One recognises immediately that a distinct difference exists between regular networks with even and with odd N. In the case of odd N a closed path which goes through all nodes of the regular network only once in one direction always exists along one type of branches e.g. in 15 b. Such a closed path which goes through all nodes only once in one direction (but not necessarily along all branches) is called a Hamilton line. Thus a regular graph with an odd number of nodes and hence nodes of even rank is composed of Hamilton lines. Since one can go along all branch lines in a closed path by following serially the Hamilton lines in the network one can state that a closed path exists which goes through all branches only once in one direction. The existence of this path called the Euler line depends on the fact that a network with nodes of even rank can be thought of as being composed of elementary graphs of second degree. Such a network or graph is then called a Euler graph. In the case of a closed path which goes through all nodes of the network only once in one direction a Hamilton line may exist along sequentially paired branch e.g. alternate ——— and —— in fig. 14b, or may not exist as in fig. 14d where any pair of different coded branches when alternated yields two nonintersecting loops through 6 nodes each as illustrated in figs. 14d1 and 14d2.



Fig. 14d1

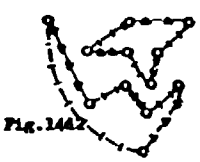


Fig. 14d2

Similarly two networks having exactly the same number of nodes and branches can be topologically different in the sense that in one a Hamilton line exists while in the other a Hamilton line does not exist. For example ten nodes of rank three each could be arranged in the forms of the following configurations, Fig. 16a and 16b.

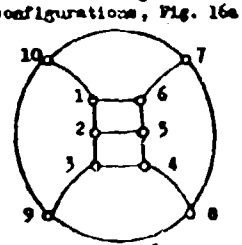


Fig. 16a

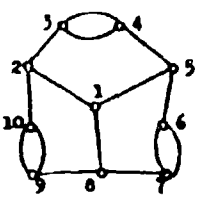


Fig. 16b

The network of fig. 16a contains closed paths through all nodes in the net such that proceeding along one of these paths all nodes are passed only once in one direction, i.e a Hamilton line. For example the Hamilton line determined by the closed path going through the nodes 1-2-3-4-5-6-7-8-9-10-1 One recognizes immediately that in the network of fig.16b the formation of a closed path which goes through all nodes only once is impossible. For example in the closed path 1-5-6-7-8-10-2-3-4-5-1 the branch between nodes 5 and 1 is used twice, forward and backward,(a shuttle) node 5 being passed twice to attain a closed path.

On the other hand taking the networks of figs. 14a, b, c and 15 a, b, one can consider the graph, i.e. networks, to consist of factors if one defines the factors of the graph as subgraphs which when taken in pairs do not have common branches, but share the same nodes. The process of analysing a graph in terms of subgraphs subject to the above condition is called factorising. Factorization of regular graphs with an even number of nodes yields elementary graphs of first degree as factors, while regular graphs with an odd number of nodes yields factors of second degree in the form of loops. (Hamilton lines and Euler lines.) In this way one arrives again at the elementary building blocks of networks.

A special case is the complete network with a maximum index of connectivity and an even number of nodes $N_K = 2s$, $\nu = (N-1)$ which consists of $2s - 1$ factors of first degree. The process of factorization of a complete network with an even number of nodes corresponds to combining $N = 2s$ elements into s pairs such that every element is paired $2s - 1$ with every other element once. For example in the complete network shown below in fig. 17a, b, c, with six nodes there are five factors shown as coded lines.

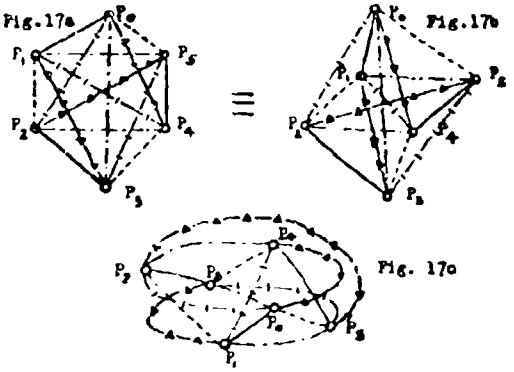


Fig. 17c

Figures 17a, b, c represent topologically identical network configurations. Geometrically 17a is the planar hexagon version, 17b the three dimensional octahedron version and in 17c the nodes are shown to be located along two parallel circumferences of a cylinder, i.e. the double belt version of 17a.

Not every graph can be split into regular factors, such graphs are then called primitive analogous to the prime numbers which defy factorization. As an example the graph of fig. 18 is a primitive graph defying factorization in as much as node C is not contained in any one of the circles of the graph.

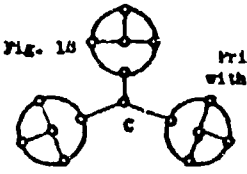


Fig. 18 Primitive regular graphs with nodes of third degree.

F. WEIGHTED NETWORKS.

A general network shall be defined as one whose nodes are of rank equal to two or larger than two. A general network must contain either all or some of the characteristics associated with trees and regular graphs discussed before, in as much as these configurations are subgraphs of the network graph. For example the network of fig.5 on page 3 is a general graph containing nodes of even and odd rank (degree). As the number of nodes of odd degree must be even in accordance with formula (2) on page 3 their rank can be increased to an even figure by addition of branches which interconnect the odd degree (rank) nodes only. Thus by insertion of four dashed line branches into the network of fig. 5 one obtains the network shown below as fig. 19. This network has now only nodes of even rank and is therefore composed of elementary graphs of second degree and hence an Euler graph as discussed before. The Euler line that is the closed path which goes through all nodes at least once and covers all branches is then the following

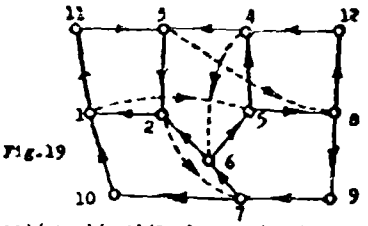
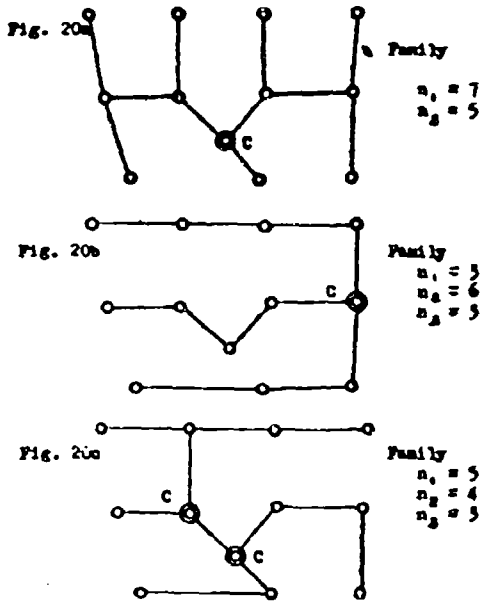


Fig.19

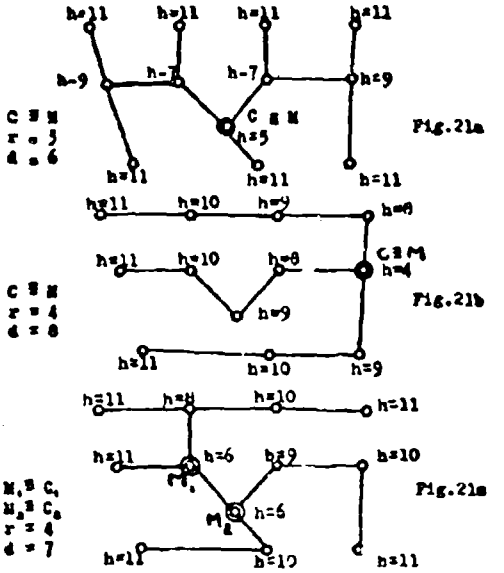
(1-11)(11-3)(3-2)(2-1) 1-5 (5-4)(4-3) 3-8 (8-12) (12-4) 4-6 (6-5)(5-8)(8-9)(9-7)(7-6)(6-2) 2-7 (7-10)(10-1)

By removing the previously inserted four dashed branches again the Euler line is broken up into four open noninterfering routes while the original network of figure 5 which contained $2p - 8$ nodes of odd degree is restored. Thus one can deduce that this method permits the establishment of p noninterfering routes (noninterfering means that there are no branches in any one of these open routes which are shared by any other of these routes) in any general network containing 2p nodes of odd rank.

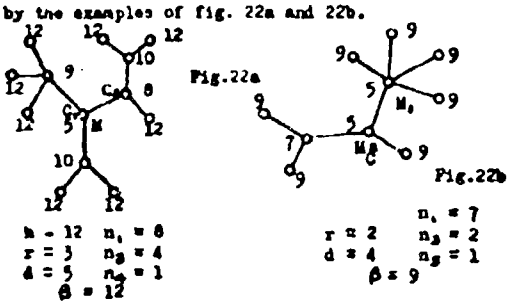
Any general network can be considered to be an aggregate of fundamental structures or trees of different family members which are superimposed on each other such that their nodes coincide. This is illustrated for the case of fig. 5 and fig. 19 below in fig. 20a, b, and c.



It has been shown before that the number of significant tree configurations (fundamental structures) becomes quite large for small numbers of nodes. For selecting a practical meaningful set of manageable size out of the possible number of significant fundamental structures and resultant distinctive networks, the topological parameters of diameter and radius of a tree and the height of its nodes in terms of branches is introduced as units. Thus one may ask for the longest path i.e. the diameter through the tree in terms of number of branches between nodes and the location of a node or nodes in the tree whose maximum distance from other nodes become a minimum. These nodes are then the centers of linear extension of the tree from where all other nodes can be reached along a path which contains at the most r branches, equivalent to the radius of the tree. The centers of linear extension and the radius of the tree are found successively by cutting off nodes of degree one from the tree, i.e. the outer nodes of the tree. In fig. 20a three cuts reduce the tree to one center of linear extension C , while in fig. 20b four cuts are required. If r is the number of cuts, the diameter of these trees is then $d = 2r = 6$ and 8 branches long. In fig. 20c however, $r = 4$ cuts reduce the tree to two centers of linear extension C_1 and C_2 and hence r - 1 cuts reduce the tree to an elementary graph of first degree whose end points C_1 and C_2 are reconnected by the axes of linear extension of the tree. The diameter of this tree in fig. 20c which has a bicenter is $d = 2r - 1$. If the successive cutting off of nodes of degree one leads to only one center, i.e. if the r - 1 cut leaves a star type configuration then the diameter of a tree with a single center of linear extension as shown before is $d = 2r$. Or one may investigate how many branches do emanate from a node to other nodes in the tree and what are the number of branches in these main branches. The maximum number of these branches represent then a topological parameter of this node in relation to other nodes which is called the height of the node. This is shown in fig. 21 below. Nodes having minimum heights are indicated as centers of mass M.



The topological height is a relative measure of connectivity of a node to all other nodes in the tree. That the center of mass need not coincide with the center of linear extension is demonstrated by the examples of fig. 22a and 22b.



The tree of fig.22a has a single center of mass but a bicenter of linear extension while the tree of fig. 22b has a bicenter of mass and a single center of linear extension. Once the nodes have been weighted in each tree configuration the trees can be superimposed and their topological heights of the nodes summed and averaged with respect to the number of superimposed trees. This is demonstrated for the network of fig. 23 which comes about be the superpositon of figs. 21a, b, c, with the resultant weights $W_i = \overline{\sum_j h_{ij}}$

height - h_{ij} trees - j nodes - i
of node j in tree i

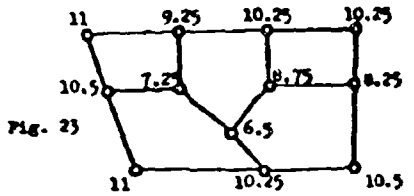


Fig. 23

These weights introduce the correspondence relations between communication system elements and the organisational relations within the larger system, e.g. the communications authority of the users associated with the nodes. Analogous to the weighting of nodes, weights can be assigned to the branches. The branch weights can be used to express the various characteristics, e.g. transmission subsystems which are symbolised by the branches independent of the weights of the nodes, or by following strictly the topological hierarchy of the nodes, the weights of a pair of nodes can be transferred to the connecting branch in terms of an average. From fig. 23 one gets then fig. 24 where the nodes are now labelled A through L.

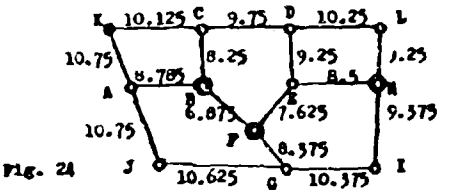


Fig. 24

The centers of mass and linear extension of the fundamental structures of figs. 21a, b, and c E, F, and H distinguished by ⊙ in the net. If one denotes by l the number of branches in a path going from one node to another and by Ω the summation of weights on the l branches in the specific path then one can evolve tables of the following type, e.g. for all possible paths which go through a node only once to link A with L, and F with D.

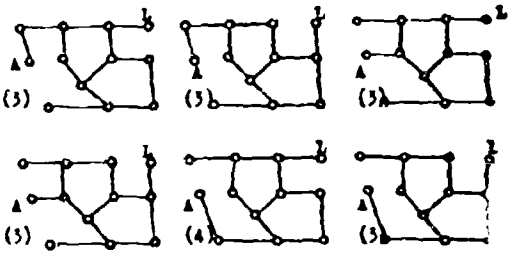
PATHS FROM A to L

Nr.of paths	Nr.of branches	Nodes passed	Nr. of centers passed	Nr. of axes passed	Ω
2	4	ABCDL	1	0	37.35
		AKCDL	0	0	40.875
3	5	ABFEL	3	1	41.035
		ABFEDL	2	1	42.850
		AJGIHL	1	0	50.375
5	6	ABFGIHL	3	1	53.035
		ABCDEHL	2	0	53.785
		AJGFEHL	2	0	55.125
		AJGFEDL	1	0	56.250
		AKEDEHL	1	0	57.625
4	7	AKCBFEHL	3	1	61.375
		AKCBFEDL	2	1	63.125
		AJGFBCDL	2	1	64.875
		AJGIHEDL	1	0	69.125
2	8	ABFGIHEDL	3	1	71.700
		AKCBFGIHL	3	1	73.375
1	9	AJGFBCDEHL	3	1	81.625
2	10	AJGIHEFBCDL	3	1	92.375
		AKCBFGIHEDL	3	1	93.250

PATHS FROM F to D

Nr.of paths	Nr. of branches	Nodes passed	Nr.of centre passed	Nr.of axes passed	
1	2	FED	1	0	16.875
1	3	FBCD	2	1	24.875
1	4	FGHLD	2	0	35.625
2	5	FBAKCD	2	1	46.285
		FGIHLD	2	0	47.625
		FGIHED	2	0	47.875
2	6	FGJABCD	1	0	56.535
		FGJAKCD	1	0	60.375
2	8	FBAJGIHED	3	1	74.535
		FBAJGIHLD	3	1	76.285
2	9	FEHIGJABCD	3	0	84.035
		FEHIGJAKCD	2	0	87.925
2	10	FBCKAJGIHED	3	1	94.875
		FBCKAJGIHLD	3	1	96.625

Noting that A can be left over 3 branches and L can be entered by 2 branches, the 3 x 2 = 6 situations for paths A to L are shown in fig.24 The number of paths for every situation is given in parentheses.



Removal of two branches out of A and one branch out of L reduces the network to a 16 - 3 = 13 branch graph for every situation with an index of connectivity of 2, equilavent to containing 2 circles and hence up to 2 x 2 = 4 possible paths at the most.

The possible situations for paths from F to D can be considered in like manner. Removal of two branches from F and two branches from D at a time in every situation yields a 16 - 4 = 12 branch graph with an index of connectivity of 1, equivalent to containing one circle, hence 2 x 1 = 2 possible paths at the most.

In general, paths between nodes X and Y of a network having N nodes and b branches, and an index of connectivity $\mu = b - \beta$ where $\beta = N - 1$ are distinguished by $\nu_x \nu_y$ situations, where ν_x is the rank of node X and ν_y is the rank of node Y. Every situation amounts to a subgraph of the network graph which contains the nodes X and Y as terminal nodes of rank one as a result of removal of the $(\nu_x - 1) + (\nu_y - 1)$ branches leading into X and Y respectively. If the original network had an index of connectivity μ then the subgraphs have an index of connectivity $\mu_s = \mu + 2 - (\nu_x + \nu_y)$ equivalent to the number of circles in the subgraph. Since there are two possible ways of passing through a circle there are up to $p_{max} = 2 \times \mu_s$ paths between X and Y in a situation corresponding to the subgraph.

D. TOPOLOGICAL IMPLICATIONS REGARDING COMMUNICATIONS SYSTEMS.

Example 1.

The tables on page 4 can be utilized as follows:

a. Statement of the problem. There are nine transmission subsystems available with which to connect a maximum of ten stations. Six of the ten stations symbolized by nodes of rank one. From the remaining four stations, two are equipped to serve as relay stations, symbolized by nodes of rank two. In general not more that five trunk groups can be terminated in a station, i.e. the highest permissible rank of a node is limited to five. In how many significantly different configurations can the available subsystems be deployed under the given conditions.

b. The solution. From the tables it follows that there are only two families of fundamental structures $(n_1 = 6,\ n_2 = 2,\ n_3 = 1,\ n_4 = 1)$ $(n_1 = 6,\ n_2 = 2,\ n_4 = 2)$ with $\sigma = 8$ significant members in the first, and $\sigma = 5$ significant members in the second family. Hence there are 13 different configurations in which to deploy the available subsystems, shown here.

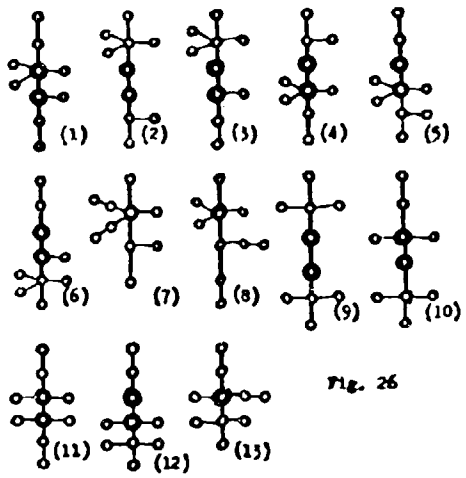


Fig. 26

Example 2.

a. Statement of the problem. The communicators associated with the ten stations of example 1 are organized along a highly centralized pattern. The layout of the communications system must conform to the centralistic structure. Thus in terms of the topological parameters discussed previously the following requirements have to be met:

Existence of only a single center of linear extension.

Minimum radius of linear extension so as to provide for closest proximity of nodes to the center node.

Coincidence of the center of linear extension with the center of mass in conformance with the organizational hierarchy of the communicators.

b. The solution. The required existence of a single center of linear extension limits the selection to graphs 7, 8, and 13 out of the available configurations of example 1. Fig. 26.

The requirement of minimum radius of linear extension is with $r = 2$ and diameter $d = 4$ branches from terminal to terminal node also met by graphs 7, 8, and 13.

The requirement of coincidence of centers of linear extension and center of mass is only met by graphs 7 and 13, in as much as graph 8 has a bicenter of mass.

Thus both tree configurations 7 and 13 will suit the requirements. A comparison of the spread of the topological heights of the nodes of 7 and 13 shows however that configuration 7 reflects a higher degree of centralization that does graph 13.

Example 3.

a. Statement to the problem. The communicators associated with the 10 stations of example 1 have different missions. Some stations may for instance serve predominantly communicators with a logistical mission. Every user organization is centralistic so far as information sharing is concerned and exchange of information between two types of user organizations usually involves the higher echelons of both organizations. The information exchange shall be accomplished easily and speedily. In terms of the topological parameters this means a requirement for:

Existence of a bicenter of mass. (One for intelligence, one for logistics)

The existence of a single center of linear extension, preferably coinciding with one center of mass and occurring close to the other.

A minimum diameter of linear extension.

A minimum radius of linear extension.

b. The solution. The only configuration which meets these requirements is that tree labelled 8, in example 1.

Example 4.

a. Statement of the problem. There is equipment available to set up six switching exchanges with four trunk groups each, where every trunk group out of the four utilizes a different type of transmission medium, or different type of transmission mode, with corresponding transmitter and receiver terminal equipment. How can the six exchanges be interconnected so that the mode or medium of the trunk groups match if each type out of the four different types of transmission modes or medium is used for a total of three transmission links between the exchanges. In terms of the topological parameters, the regular network with six nodes of rank four each has to consist of four factors where the factors are elementary graphs of first degree which when taken in pairs do not have common branches but may share common nodes of the net. Every factor must comprise three branches in the network. Which has $b = \frac{1}{2} \times 6 \times 4 = 12$ branches.

b. The solution. The regular network may first be considered as a complete network where every node is of rank five. This network is then a special case of the complete regular network with an even number of nodes $n = 6 = 2s$ where $s = 3$ consisting of $2s - 1 = 6 - 1 = 5$ factors in the form of elementary graphs of first degree, as shown in figs. 17 a, b, and c. Elimination of one factor, for instance, the one consisting of

the three branches as shown by ——— of fig. 17 reduces the complete network to the required network with nodes of rank four, in the form of the required four factors, symbolizing the four types of different transmission media or modes associated with every trunk group emanating from a switching center as shown below.

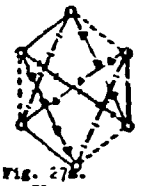


Fig. 27a. Planar hexagon version.

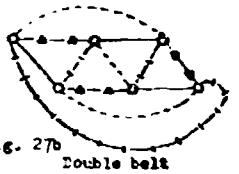


Fig. 27b Double belt version.

Example 5.

a. Statement of the problem. In the network of example 4, the status of subscribers in the local area of the exchanges, or the operational state of the exchange is subject to change. These changes have to be brought to the attention of the supervisors of all other exchanges in the form of supervisory messages containing, for instance, directory updating information, etc., such that the originator of the supervisory message is able to determine if all other exchange supervisors have received his message. The procedure for handling the supervisory messages of this type shall be such that the least amount of channel capacity in the net is diverted from subscriber service to the system supervision and control.

b. The solution. In terms of the topological parameters, supervisory messages can be circulated along a closed path going through all nodes only once in one direction, that is, through Hamilton lines which in the given case are made up of two out of the four factors, e.g. following a path out of and back into the originating node along alternating coiled lines. There are $\binom{4}{2} = 12$ variations possible hence 12 closed paths through the six nodes can be utilized. For example, the first originating supervisor may choose —a—a—, the second supervisor having received the message via —a—a— marks it —a—a—, notes the content and sends it out over - - - - -. The third supervisor receives the message over - - - - -, marks it - - - - -, and must now send it out over —a—a—. From now on the message has the predetermined routing —a—a—, - - - - - which leads back to the originator, who upon correct reception of his own message is thus insured of delivery to all others. If one of the supervisors does not want to acknowledge the content because of garbled text, he marks failure of acknowledgement and may send the message forward —a—a—, - - - - - or backward - - - - -, —a—a—. In either way the originator who gets his message back, in the first case over - - - - - when the routing has passed through all supervisors, even though some of the supervisors did not acknowledge the text, in the second case the originator gets his message back over the same trunk labelled, over which he sent the message out. Thus the supervisory message may not only serve for the exchange of supervisory messages between exchanges, but also as a test patterns, from which the originator obtains new information about the operational status and the performance of the system, with a minimum diversion of channel space for system supervision and control. In general the implications, regarding communications, of the existence of Hamilton lines in a network in the light of the forwarding and processing of supervisory messages are easily recognized by a comparison of the networks of fig. 16a and 16b. In the network of fig. 16 a the Hamilton line permits the procedures of supervision discussed before, while it is readily recognized that in the network of fig. 16b much more sophisticated procedures of supervisory message forwarding and processing are necessary. There one of the switching centers must pass the message on at least twice in two different directions accompanied by a rather tedious recognition and booking procedure to make sure that every one is notified, which in itself produces an excessive supervisory traffic loading.

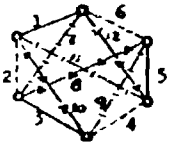
Example 6.

a. Statement of the problem. All of the trunks in the network of example 5 shall be protected by loading them with pseudo random carriers which are emanated and controlled from one master key generator which can be located anywhere in the network. To mention a controversial subject, the network shall be synchronized with one master clock located anywhere in the network. Another application would be a test procedure for the connectivity of the net by sending a test signal out and having it returned back through a closed path that goes through all trunks of the network.

b. The solution. In terms of the parameters discussed the network must contain an Euler line and hence it is required that the nodes of the network are of even rank. In the case of the networks of example 4 the Euler line is given by branches labelled 1 through 12 in fig. 28. The Euler line leads through the branches labelled consecutively 1 to 12.

Fig. 28

Euler line.



Example 7.

a. Statement of the problem. The network shown in fig. 29 has 6 nodes of rank 4 each, serving several different equally important subscriber organizations which are associated with the six exchanges A, B, C, D, E, and F. The single center of the —— organization tree is closely associated with exchange C. The radius of linear extension of the —— organization shall be two branches. The center of linear extension and the center of mass coincide. There is another - - - - - - subscriber organization which is served by the net, which has two centers of linear extension associated with D and E, and a third — · — · — subscriber organization with a single center of linear extension at D, but two centers of mass, one coinciding with D, the other coinciding with E.

The resultant weighting of the network is required.

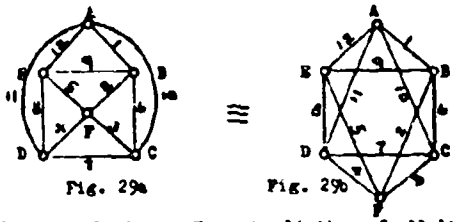


Fig. 29a

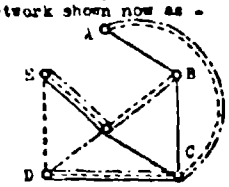
Fig. 29b

b. The solution. From the listing of all distinct trees contained in the network, fig. 15b, one picks the tree configurations labelled by the combinations 6,7,1,3,5 which fits the —— organization; the tree 7,8,10,2,5 which fits the - - - - - - - organization, and the tree 7,10,2,4,5 which fits the — · — · — organization as shown individually



6,7,1,3,5 7,8,10,2,5 7,10,2,4,5

The results of the superposition of the trees leads to the network shown now as -



Trunks between E and F/ and D - C carry information for all three organization on a first choice basis; trunks between F and B, and between A and C (branches 2 and 10) in the net have to carry information for the - - - - - - - and the — · — · — organizations. Thus the selection of the transmission media for the net must be made in such a way that the connectivity between the nodes E and F/ and D is retained with utmost reliability, since its impairment would effect all three equally important user organizations. The rest of the branches may be made less costly. In this way a hierarchy in the universality of the branches in the network relative to a human user system is developed which influences design decisions, while from the listing of all possible distinct trees in the net, other types of user organizations symbolized by trees can be found which would also suit the design.

4. CONCLUSIONS

From the limited given examples and the discussion there is indicated the importance and significance of topology engineering of communications networks for the establishment of guide lines, criteria and methods, to make a given or required set of nodes of different rank (symbolizing central office subsystems of different size), connectible with a given or required set of branches (symbolizing transmission subsystems) in the form of communications networks which have to meet conditions related to the organizational structure of the communicators, message handling procedures, system supervision, control, and operation.

REFERENCES

Lewin, K. - Principles of Topological Psychology. (Book - McGraw-Hill).

Konig - Theory der Graphen.

Shimbel - Structure in Communications Nets. (Symposium on Information Networks, Polytechnic Institute of Brooklyn, April 12-12, 1954).

Rademeister - Einfuhrung in die kombinatorische Topologie. (Book)

Gilbert, E. N. - Enumeration of Labelled Graphs. (Canadian Journal of Mathematics. (Vol. 8, No. 3 pp 405-411, 1956).

SERVICEABILITY: COMPLEMENT TO RELIABILITY

By: Mr. R. E. Wilcox and Cdr. V. R. Wanner
Information Systems Branch, Office of Naval Research

INTRODUCTION

The total lifetime of equipment may be thought of as the summation of its "up" times and "down" times between the extremes of assembly line and junk heap. A perusal of the literature shows that over the years the predominant current of our efforts in reliability has been toward maximizing the up times. This is as it should be. However, even in our latest hardware, when we string together an increasing number of critical elements (albeit with smaller individual probabilities of failure) the overall probability of failure remains significant and the down times are still with us. Although in many systems (e.g., missiles and satellites) nothing can be done once failure occurs, there still remain a great many complex systems in which equipment malfunctions are repaired and the gear is returned to service.

This subject is usually treated under the heading of "maintainability." However, that term refers properly to the maximizing of up time for each unit period of down time (1). We therefore have chosen the term serviceability for discussion in order to emphasize the importance--and the neglect-- of that property of complex equipment which, conceding the likelihood of periods of sub-optimal performance (including complete failure), permits us to minimize the time during which the system is out of service. Specifically, the word serviceability will include here the capacity of a system for permitting ready recognition of sub-optimal performance, simple diagnosis of the cause of trouble, easy rectification of the difficulties, and rapid re-optimizing of performance. The relationship between these factors is illustrated in Fig. 1.

In this paper we have two objectives:

(a) To establish the status of serviceability as an important complement to the more conventional reliability considerations in the overall problem of equipment dependability.

(b) To propose that a theoretical approach to serviceability would be helpful in delineating and analyzing the necessary experiments and intuitive design and specification work.

THE IMPORTANCE OF SERVICEABILITY

During periods of outage of vital equipment, a military vehicle of war often moves from the asset column in national defense to the liability column--the warship becomes an ostentatious ferryboat and the aircraft an expensive (and probably expended) viewing platform. It was mentioned above that serviceability includes the time required to discover sub-optimum performance; in operational military systems this period is particularly dangerous because the onset of difficulties is not always obvious. At such times the military commander assumes capabilities which he does not have, and the consequences are attested by myriad groundings, collisions, and aircraft crashes, without even considering missed opportunities in time of war. Anyone who has worked with radar, sonar, or communications gear has experienced the uncertainty of wondering whether the other fellow really isn't there or whether the gear just isn't up to snuff.

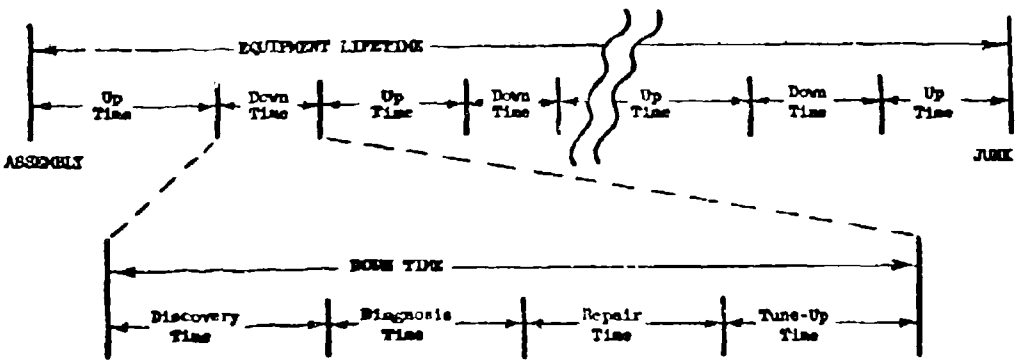


FIGURE 1

There is another problem which is affected by serviceability, that of personnel. For years the military departments have been fighting a continuous battle to supply the required quantity and quality of trained personnel on site with equipment. This has been an expensive sink for manpower, time, and money, and the problem is increasing not only because of growing equipment complexity but also because of the increasing number and variety of functions which are being automated and mechanized. We are approaching an impasse wherein everything requires above average personnel! Improved serviceability would help in two ways. First, shorter checking and servicing times would enable the available competent technicians to maintain more pieces of equipment (or require fewer personnel for the same equipment). And second, simplified maintenance procedures would permit the utilization of less-well qualified men. Note that these advantages not only decrease the direct maintenance problems but also reduce the dollar, time, and manpower costs of training.

THE NEED FOR ADDITIONAL EFFORT

To the engineer who has devoted his professional lifetime to minimizing the potential down time of military hardware, it may seem trite for users to urge that greater effort is required. Certainly the point has been argued many times before. However, a suspicion that certain facets of the problem have received short shrift is confirmed by an examination of the literature.

We have applied much of our science and ingenuity to the development of equipment which exhibits a maximum time between failures under operating environments. The need for this has been established through careful studies (2) and is reflected in government contract specifications (3). Various approaches are used (4), including improved dependability of individual components, redundancy both in the form of circuitry and as spare parts, and underrating of equipment. In addition, a very considerable body of theory has been built up to delineate desirable tests and avenues of improvement, enable measurement and analysis of experimental results, and permit prediction of operational reliability (5,6,7). Such a large effort devoted to increased mean time between failures is certainly justified and highly desirable, and has proven itself worthwhile.

It is also true that some effort has been devoted to the problems of maintenance and servicing. Manufacturers deserve considerable credit for their accomplishments in making component parts accessible, numbering of components, and color coding of wiring. The instruction books and operator's manuals supplied with hardware offer extremely detailed procedures, effective illustrations, and illuminating descriptions. In many cases well designed special test equipment is provided. However, the extent to which such aids to serviceability are provided is left primarily to the conscience and ingenuity of the system designer. The results are unsatisfactory, as noted in the preceding section. This is not in any way a criticism of either the skill or the ethics of our systems designers; just the opposite is intended--such admirable effort is deserving of the theoretical and experimental backing which could make it many times more effective.

Many appeals have been made in the past for improved serviceability (7,8,9), but appear to have gone largely unheeded. The reasons underlying this are rooted in two predominating philosophies found in design, marketing, and procurement. First, system designers and specification writers assume implicitly that a level of competence approaching their own will be available on site for equipment maintenance; this assumption is almost never justified in practice. Second, built-in improvements to serviceability are expensive initially, and the immediate solution to rising costs on the one hand and maintaining a competitive position on the other is to set aside such serviceability "frills." Unfortunately, the result is simply to pass the problem and its ultimate cost to the user.

At first glance, the solution appears to lie with the writers of procurement specifications. If "Mil Specs" called for given levels of serviceability, suppliers could bring the full power of their R and D groups to the problem without fear of jeopardizing competitive position, knowing that all other bidders were facing the same requirements. But a more careful look shows that the necessary tools are not available to the specification writers. They must know what improvements are feasible and desirable, and--even more important--they must be able to state the resulting specifications quantitatively.

POSSIBLE AVENUES OF IMPROVEMENT

It was noted in the Introduction that serviceability included improvements to four components of down time: recognition of suboptimal performance, diagnosis and localization of the cause of trouble, rectification of the difficulties, and re-optimization of the system. We will briefly examine each in turn.

The time required to establish the presence of a malfunction varies from the instantaneous (in the event of equipment "blow up") to much longer periods--hours, days, even weeks in the case of the gray area of diminution of performance between regularly scheduled tests and checks. The ideal, of course, is a built-in continuous monitor which will provide an immediate alert when performance falls below prescribed

standards--an operational analog of the "tilt" light in pin-ball machines. Actually, such monitors could indicate not only satisfactory overall performance but also the proper functioning of the various system components, thus providing a measure of self analysis of troubles. Even when trouble is known to be present, excessive down time may result from difficulty in diagnosis. Here again the time required is a function of the nature of the casualty; obvious failures (such as charred parts) are located with a quick visual inspection, while obscure performance degeneration or intermittent irregularity require much longer periods. In military equipment, trouble shooting is frequently a point-to-point probing process, hampered not only by high packing density and complex circuitry but also by constraints of the physical space within which equipment is located. Rapidity of diagnosis is a direct function of personnel competence. One possible solution is the continuous component monitors mentioned above; another is diagnostic programming as practiced on many general purpose digital computers.

In the case of making repairs, it is generally assumed that every red-blooded American boy knows how to fix things. This belief is not entirely unjustified; to a considerable extent it is just such innate know-how that keeps our contemporary equipment running as well as it does. But such capability is by no means universal, and if the military technical manpower problems are to be alleviated then hardware must be designed to require no more skill than that necessary to unplug and replace modules; tweezer operations on miniaturized circuits and good solder joints must not be required for operational maintenance.

Even when skillful repairs have followed correct diagnosis, equipment frequently is not returned to full design performance because of difficulties in "tune up," or re-optimizing the system for harmonious operation with the new or modified components. Here the objectives are a minimum number of sub-systems to be re-optimized, relatively few, uncomplicated steps involved in the re-optimization, a minimum of auxiliary test equipment, accessibility of adjustments, and lack of excessive sensitivity.

THE VALUE OF A THEORETICAL APPROACH

It may well be inferred from the preceding comments that serviceability can be bought only at a price. Our approach thus far has been intuitive; valid as the arguments may be, no-one is willing to put out hard cash until he knows the worth or the effect of his purchase. As stated above, specification writers must be able to state their requirements quantitatively.

The situation may be expressed as a sort of "boundary" problem. On the one extreme is equipment characterized by low initial cost but poor serviceability--a jungle of components and interconnecting wires whose maintenance places a premium on personnel competence. It contains no integral performance-monitoring indicators, no inherent aids to diagnosis of trouble, and requires much removal of interference and trial-and-error activity in repair. Further, tuning up is complicated by a multitude of sensitive, inter-acting variable resistors and tuned circuits which require hours of adjustment and readjustment in conjunction with much auxiliary gear. At the other extreme lies a highly automated system characterized by high initial cost but serviceability features limited only by the state of the art--a well organized collection of color-coded encapsulated modules requiring only replacement of an automatically identified plug-in unit when trouble occurs. Multiple circuit redundancy provides extreme reliability and permits module replacement without interrupting operations (10). All modules are factory pre-tuned and pre-adjusted, but the system is self-optimizing or self-adapting to specific situations.

To specify a system for any given application it is necessary to select an optimum position between these two boundaries; optimum in the sense of balancing requirements and desirable servicing aids against cost. Note that cost here means not only initial price but also the men and dollars required for maintenance over the lifetime of the equipment. This calls for a measure of serviceability, one sufficiently detailed to account for the various factors involved and designed for use with actual systems or experiments (i.e., the parameters must be physically meaningful and determinable). Such a measure could be used by the systems designer to identify the contributing parameters and the effect of varying them in his system. Obviously it could be used also by specification writers in describing desired equipments. But it could also be used by circuit engineers, psychologists, field maintenance specialists, and operations evaluation personnel to determine what variables to study and the value of various design approaches. And perhaps most valuable, the operations researchers could apply queuing theory and programming theory to establish optimum compromises between the parameters (11). At the very least, such a measure would provide a means for comparison of alternative systems.

To illustrate the principles of such a measure, an elementary figure of merit will be presented here. It is not considered to be the final answer, nor is the approach particularly original. But it should serve to identify some of the major parameters, demonstrate the relationship

of serviceability to more traditional aspects of reliability, and suggest one means of analysis. In this approach time serves as the basic standard of measurement and statistics provides the appropriate methods of combination.

A reasonable figure of merit for overall systems dependability is the ratio of satisfactory operating time to total time of intended operation:

$$M = \frac{T_s}{T_t} = \frac{T_t - T_d}{T_t} = 1 - \frac{T_d}{T_t}, \qquad 0 \leq M \leq 1, \tag{1}$$

where

M = figure of merit,

T_s = time during which operation is satisfactory

T_t = total time in commission, and

T_d = down time.

A more detailed examination of down time shows it to be a function of four parameters:

$$T_d = f(t_c, t_d, t_r, t_h), \tag{2}$$

where

t_c = checking time, or time to discover the presence of trouble,

t_d = time to diagnose and localize the trouble,

t_r = time to repair the trouble, and

t_h = time to tune-up or "reharmonize" the system.

These must all be present whenever a malfunction occurs and is remedied, although one or more of them may be essentially zero (e.g., when occurrence of the trouble is instantaneously evident). On the other hand, checking or testing may be performed when no trouble is known to be present, and thus it must be treated separately from the other three. Hence we break down checking time further into time spent in normal or regular performance checks (t_{cn}) and time spent in special checks when trouble is suspected but may or may not be present (t_{cs}).

The time actually expended in the various categories will vary between individual instances, and thus must be treated statistically. We define $p(t_i)$ as the probability that the time spent in the i-th category equals t_i (e.g., $p(t_r)$ is the probability that the time required to repair a given trouble is t_r minutes). If we assume that time is quantized into units (e.g., minutes) which are arbitrary but identical for all categories, and further that time spent in any category is independent of that spent in the others, then we may write

$$p(T_d) = \sum_{cs} \sum_{d} \sum_{r} \sum_{h} \delta(T_d, t_j)\; p(t_{cs})\; p(t_d)\; p(t_r)\; p(t_h) \tag{3}$$

where

$$\delta(T_d, t_j) = \begin{cases} 1 \text{ for } T_d = t_{cn} + t_{cs} + t_d + t_r + t_h \\ 0 \text{ elsewhere.} \end{cases} \tag{4}$$

We can now define mean time to recovery $\overline{T}_d$ as

$$\overline{T}_d = \sum_{d} T_d\; p(T_d). \tag{5}$$

Returning to the figure of merit of Eq. (1), we now have a means of determining the contribution of each of the categories of down time to the overall condition of system reliability. But we can go one step further. Conventional reliability theory gives us a quantity called "mean time between failures." We might use this as $\overline{T}_s$, the mean time during which operation is satisfactory. Although the derivation will no longer be physically rigorous, we may still retain a figure of merit which is useful, intuitively satisfactory, and much simpler to obtain than one which is statistically unquestionable. In these terms we have

$$M = 1 - \frac{\overline{T}_d}{\overline{T}_s + \overline{T}_d} \tag{6}$$

which can be obtained entirely by experimental measurements or theoretical calculations. Again it is emphasized that the approach given here is intentionally oversimplified. For example, the assumption of independence of time spent in the various categories of down time is questionable. However, two important points are demonstrated: standard statistical techniques are adequate to handle the categorizations and inter-relationships required, and theoretical approaches are feasible which are also compatible with existing work on reliability theory.

Of course, a realistic solution to the serviceability "boundary" problem must be more sophisticated. Parameters of the distributions describing the variables would have to be evaluated experimentally, and a theory should be constructed both to provide an understanding of the behavior of the variables and to permit prediction of their values in proposed systems. In any given system design problem, various values of the figure of merit might be related to dollars (both initial and maintenance costs) and a curve drawn to establish the optimum trade-off. However, in real life there exist other less tangible considerations which cannot be ignored, such as currently available funds, equipment priorities, and personnel implications.

CONCLUSIONS

Experience demonstrates repeatedly that much remains to be done in improving the serviceability of complex military systems and equipment. The method presented here serves as an elementary example of the theoretical approach to isolation and analysis of the factors involved in serviceability. Hopefully, it will point the way toward means of achieving more reasonable trade-offs between initial equipment cost and subsequent down time, and of establishing an economically sound basis for partially substituting equipment capabilities for highly skilled maintenance personnel on site.

It may have appeared that this is a user's plea for improved serviceability, and so it is. Nevertheless, we submit that there is something in serviceability for those of you on the other side of the fence. For procurement personnel there is the possibility of getting more per dollar and the satisfaction of providing more dependable equipment for military operations. And for the producer there is not only the incentive for increased sales but also the pride of hearing his customers say, "this is a damn fine piece of gear!"

References

1. "Glossary and Dictionary of Terms and Definitions Relating Specifically to Reliability," C. M. Ryerson, ed., Proc. of 3rd National Symposium for Reliability and Quality Control in Electronics, pp. 59-84, Jan. 1957.

2. "Reliability of Military Electronics Equipment -- 4 June 1957," Report by the Advisory Group on Reliability of Electronics Equipment (OASD-R&E).

3. Bucci, E. J., "The Navy Specification Program for Reliability," IRE Trans. on Reliability and Quality Control, Vol. RQC-14, pp. 27-32, Sept. 1958.

4. Fein, Louis, "Insurance Systems for the Management and Control of Error," Final Report to the Office of Naval Research under contract Nonr-2554(00), Jan. 1960.

5. Moskowitz, F., and McLean, J.B., "Some Reliability Aspects of Systems Design," IRE Trans. on Reliability and Quality Control, Vol. RQC-8, pp. 7-35, Sept. 1956.

6. Youtcheff, J. S., "Statistical Aspects of Reliability in Systems Development," IRE Trans. on Reliability and Quality Control, Vol. RQC-12, pp. 25-36, Nov. 1957.

7. Whitaker, G. C. F., "The Problems of Reliability and Maintenance in Very Large Electronic Systems for Shipboard Use," Journ. Brit. I.R.E., Vol. 19, pp. 625-646, Oct. 1959.

8. Kirkman, R. A., "Field-Testing Reliability into a Complex Weapon System," Missile Design and Development, Vol. 6, No. 12, pp. 20-23, Dec. 1959.

9. Smigh, J. R., "Reliability and Maintenance Design Considerations," Proc. of 3rd National Symposium for Reliability and Quality Control in Electronics, pp. 264-269, Jan. 1957.

10. Wasserman, R., et al, "Improvement of Electronic Computer Reliability Through the Use of Redundancy," Hermes Electronics Co. Publication #4-828, April 1960.

11. Heyne, J. B., "Integrating Reliability Considerations Into Systems Analyses," Proc. Western Electronic Show and Convention, Part 6, pp. 30-31, Aug. 1958.

RELIABILITY ANATOMY for SYSTEM DESIGN ENGINEERS*

E. S. Winlund, Manager, Systems Administration
Stromberg-Carlson San Diego

Abstract: In order to design systems satisfactory to military and industrial customers, with economy of design effort and production cost, it is desirable to (a) determine achievable system reliability, as well as performance and price, and (b) integrate system performance and reliability indices with long-term users cost, to achieve optimal system design. This paper develops a straightforward technique for (a) as groundwork for a subsequent paper on (b) which develops a system economy model from the users viewpoint. It is done in design-engineers language, rather than in traditional statistical terminology.

The system user, the customer, is concerned only with (a) the performance capability for a needed function, (b) the reliability or availability of that performance when needed, and (c) the economy of the system with respect to justifiable cost. To the producer, therefore, these are the three product assurance considerations.

When hourly user costs are assigned to maintenance and unreliability delay times in a simple mathematical model of the system, there is an engineering management dividend. The incremental user costs resulting from mathematical insertion of redundancy into the system components (at all system levels) throws a bright spotlight on relative economy of design effort in the various areas.

DEFINITIONS of terms, as used in this and subsequent papers, follow:

Reliability: Generic and qualitative usage only; not survival.

Survival: Probability of no failure during the given task time. Failure: Independent occurence causing unacceptable output. Failure Rate: Mean number of failures per stressed hour, or per million hours for parts. Reciprocal of Stress Life.

Stress: Electrical, mechanical, thermal, or other forces contributing to failure. Stress Life: Mean stressed time between failures, in hours; "MTBF"

System: Complete operational assembly, including "situation". Hardware System: Full complement of hardware components. Situation System: All other system components, including personnel, environment, and inputs.

Component: Generic for any constituent of the system at any level. Package: Assembly of parts. Part: Basic component such as transistor or bearing.

RELIABILITY INDICES: In order to establish a common time basis for reliability definitions and calculations, figure 1 shows a time allocation made up of the following elements:

Off time is that when the component is shut down because there is no need for it. Standby time is that during which the component is not in use, but nevertheless it has not been shut down because further use is anticipated.

Prevention time is that part of regularly-scheduled maintenance during which the component is otherwise operable and could soon be returned to operation upon demand.

Down time is that during which the component is not operable because of a failure. It includes all unscheduled maintenance time plus that part of scheduled maintenance time during which the component is not operable due to a failure.

Rerun time is that during which the component is operating properly but is occupied rerunning a task because of a failure. If it occurs it normally follows a period of down time, but not necessarily. For example a human component may make a mistake, causing no down time, yet rerun time is required. Rerun time can be a sizeable item in data processing systems, but in many systems it does not occur; if not, it is set equal to zero in any equations and they remain valid.

Delay time is the sum of down and rerun time. That is, it is the total lost time due to failure

*This is a 3:1 abridgement of a December 1959 report of the same title and author, then as Consultant, Computer Department, General Electric Co. Copies available from the author.

TIME ALLOCATION

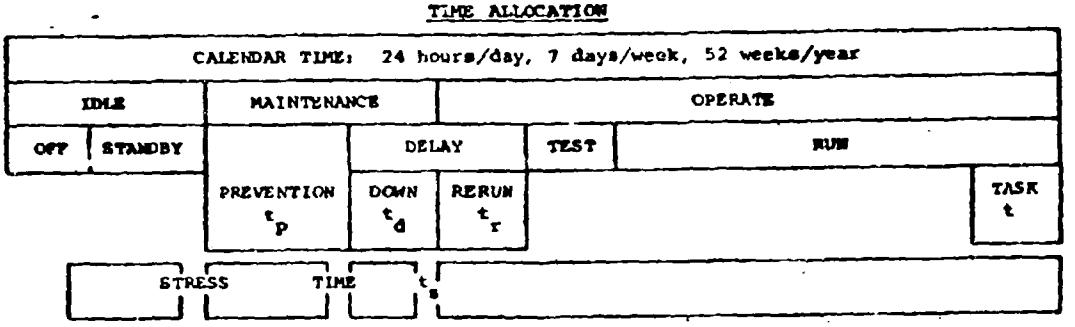


Figure 1

Test time may be considered as that conducted by the operating crew for testing programs, etc. Run time is that used productively for successive tasks. Test and Run time are seldom used in the reliability calculations, however.

Task time is that required to accomplish a specific task, assuming no trouble. It is used primarily in the calculation of survival, or probability that the task will be completed without failure.

Stress time is the total time the component is undergoing the kind of stress that contributes to ultimate failure. In most electronic equipment this is merely the time that power is applied. For mechanical devices it is usually the time they are in motion. For structures it may be the time they are undergoing vibration.

In the equations to follow it is assumed that a component is undergoing stress most of the time it is in maintenance, because it is either powered or is subject to unusual stresses such as "flying screwdrivers". Most electronic equipment is assumed to be under stress during its standby periods, and some types of equipment can be under stress during part of their off periods.

THE INDICES

$$\text{Stress Life } T_f = \frac{\text{Stress time } t_s}{\text{Number of failures } f}$$

This is the mean stressed time between failures, or "MTBF". Its true value is estimated from the long-term accumulated stress time on one or more identical components, divided by the accumulated number of failures that occured therein during the stress time.

$$\text{Prevention/Stress } U_p = \frac{\text{Prevention time } t_p}{\text{Stress time } t_s}$$

This is the fraction of stress time spent on preventive maintenance, obtained from long-term maintenance records.

$$\text{Down/Stress } U_d = \frac{\text{Down time } t_d}{\text{Stress time } t_s}$$

This is the fraction of stress time that the component is inoperable due to failure, obtained from operating and/or maintenance records.

$$\text{Rerun/Stress } U_r = \frac{\text{Rerun time } t_r}{\text{Stress time } t_s}$$

This is the fraction of stress time that the component is operable but occupied in rerun due to failure, obtained from the operators log.

The above three indices are actually the fractional unavailability times (hence the U symbol)but names like "prevention unavailability" are so awkward that the above are considered simpler and more informative.

Failure Rate $\lambda = 1/T_f$. The usefulness of this index is greatest at the part and package level. Expressing it in failures per million hours provides numbers easier to handle and less subject to error than percent per thousand hours.

Maintenance/Stress $U_m = U_p + U_d$

Maintenance/Failure $T_m = T_f U_m$

Delay/Stress $U_y = U_d + U_r$

Delay/Failure $T_y = T_f U_y$

The primary reliability consideration for many, if not most, systems is "availability" rather than stress life or survival for a particular task time. On the other hand survival is the primary consideration for some systems, such as missiles, beyond the time when maintenance stops. For many systems both must be taken into account. We define "availability" as the proportion of total stress time normally available for productive use of the component.

Continuous Availability

$$A_c = 1 - U_p - U_d - U_r$$

This is the long-term fraction of stress time that the component is available for productive operation, taking into account all maintenance time and any rerun time typically required for the specific application. If there is no rerun, $U_r = 0$.

Continuous availability also may be regarded as the probability that the component will start proper operation, for random trials at any point in calendar time. Thus it may be handled with probability math.

Demand Availability $A_d = 1 - U_d - U_r$

This is the fraction of stress time that the component can be made available upon demand, pulling it out of preventive maintenance if necessary. It is stress time less delay time. It is a short-term consideration, because preventive maintenance usually cannot be postponed very long without consequences.

Survival $A_t = e^{-t/T_f} \approx 1 - t/T_f$

This is the probability of completion without failure of a task requiring time t. It is often called the "reliability" when maintenance and delay are not considered. It can be considered the availability of completed tasks, assuming successful starts.

THE PRODUCT RULE

"System or group Availability of components in 'series' is the PRODUCT of all component availabilities"

One often hears that the above product rule is "pessimistic" or "overconservative". This is about like saying algebra or logarithms give a slightly wrong answer. If the conditions on which the product rule is based are fulfilled, and we believe in the logic of mathematics, the product rule has to be right. Here are some of the commonly-ignored conditions leading to this sad reputation. TO AVOID PESSIMISTIC RESULTS:

1. Stress life must be related to STRESS TIME, never to calendar time.
2. Mechanical assembly life must be related to MECHANICAL STRESS TIME or actuation, not to electrical stress time.
3. Only INDEPENDENT failures may be used to calculate Stress Life, never dependent failures nor replacements.
4. EXCLUDE or de-weight components whose failure will NOT cause unacceptable system performance.
5. SYSTEM down time resulting from component failure, NOT component down time, must be used if the replaced component is repaired later.
6. Delay must be calculated by PRODUCT RULE if accuracy is desired. Addition yields a too-high result.

REDUNDANCY

The probability that a component will be available, and the probability that it will not, must add to unity. Similarly the sum of Survival and probability of failure is unity. Let a_1 be the Availability and u_1 the Unavailability of component #1, with a_2 and u_2 for component #2. Then the binomial expansion gives:

$$a_1 + u_1 = 1 = a_2 + u_2 = a_3 + u_3$$

$$(a_1 + u_1)(a_2 + u_2) = 1$$

$$= a_1a_2 + a_1u_2 + a_2u_1 + u_1u_2$$

a_1a_2 is the simultaneous Availability

a_1u_2 is Availability of #1 when #2 is not.

a_2u_1 is Availability of #2 when #1 is not.

u_1u_2 is the simultaneous Unavailability.

For the Availability of at least one or both, we need only drop the last expansion term. The procedure may be used for any number of components. For the special case of two "equal" components (equal Availability or Survival) we note that the simultaneous Availability is a^2 and "one or the other" is $2a(1-a)$. Thus total Availability is $2a-a^2$.

Availability of two unequal components:

$$a_g = a_1a_2 + a_1(1 - a_2) + a_2(1 - a_1)$$
$$= a_1 + a_2 - a_1a_2 \quad \text{total.}$$

Availability of two equal components:

$$a_g = a^2 + 2a(1 - a) = 2a - a^2 \quad \text{total.}$$

Availability of three equal components:

$$a_g = a^3 + 3a^2(1 - a) + 3a(1 - a)^2$$
$$= 3a - 3a^2 + a^3 \quad \text{total.}$$

SPARE REDUNDANCY is the use of an unstressed component or two to back up the stressed component performing a task. It can be used at any level of the system. Upon failure of the stressed component, the spare is assumed to immediately take its place under stress. This in turn implies immediate failure detection, automatic or quick manual switching, high reliability of the detection and switching system, and repair of the failure before the next task starts (if the redundant Stress Life gain is wanted immediately).

For analysis we can say that the spare component (#2) effective failure rate λ_s is 0 until switched in and stressed. Its probability of being switched in is the probability of failure of the initially-stressed component #1. So we can write:

$$\lambda_s = \lambda_2(1 - e^{-\lambda_1 t})$$

Group (stressed #1, spare #2) Availability is: $e^{-\lambda_g t} = a_g = a_1 + a_2 - a_1a_2$

where $a_1 = e^{-\lambda_1 t}$ and $a_2 = e^{-\lambda_s t}$

Then we have Stress Life gain as

$$T_{fg}/T_f = \lambda_1 t / \lambda_g t$$

which may be evaluated from the above.

STRESS LIFE GAIN: It is usually easier for the engineering designer to picture Stress Life rather than Survival. Stress Life gain, or the ratio of redundant to component Stress Life, is plotted in figure 2. Also the advantage of using "spare redundancy", rather than "stressed redundancy", is shown.

For a single component the Stress Life gain of course remains at unity regardless of task time, as shown by the line along the abscissa. For two equal stressed components the lower curve applies. For three equal stressed components, or for one stressed and one equal spare component the central almost identical curves apply. For one stressed and two equal spare components

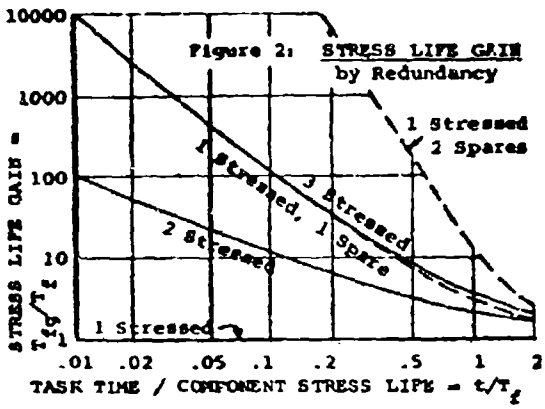


Figure 2: STRESS LIFE GAIN by Redundancy

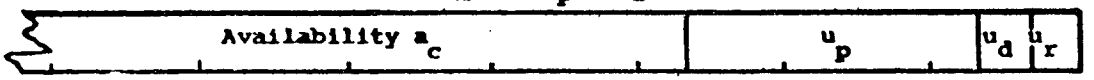
the top curve applies. Obviously stressed redundancy can provide remarkable Stress Life gains, and spare redundancy can add orders of magnitude to these gains.

But these are not unmixed blessings. Note that the Stress Life gain, and therefore reliability, is very sensitive to task time. The more redundancy the more sensitive. In fact as task time approaches the component Stress Life, the redundant advantage is about lost.

In most systems the task time is extremely small in relation to part life, and therefore redundancy applied at the part and package level may be expected to approach these gains. At the major component level in complex systems, on the other hand, task time frequently runs 10 percent or more of major component Stress Life. In such situations the Stress Life gain in adding a redundant major component may not be enough to justify its cost.

REDUNDANT RELIABILITY MODELS: Since spare redundancy may be attractive for the achievement of adequate reliability, we will look at it more closely with respect to single- and double-stressed redundancy. Following are three simple reliability models, or sets of equations, for the significant relationships. Here we introduce a numerical example in order to visualize the effect.

Assume	One Stressed Component		Example
u_p = .0998	Continuous Availability	$a_c = 1 - u_p - u_d - u_r$	= .8663
u_d = .0184	Delay / Stress	$u_y = u_d + u_r$	= .0339
u_r = .0155	Maintenance / Stress	$u_m = u_p + u_d$	= .1182
T_f = 32.5			

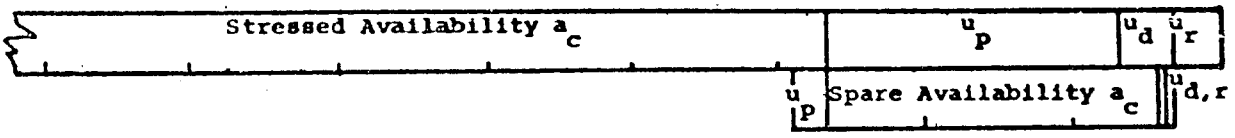


Stress time = .65 .70 .75 .80 .85 .90 .95 1.00

Continuous Availability is unity minus Prevention/Stress, Down/Stress, and Rerun/Stress time ratios. The 86.6% is obtained from the three unavailabilities listed at the left, which are actual major component group figures taken from certain installations. Delay/Stress and Maintenance/Stress are similarly calculated. The Availability breakdown is shown on the bar graph from 65 to 100%.

Now proceed to text to the right.

One Stressed and One Equal Spare Component: Lets consider precisely how the spare might be used, referring to the Availability bar-graph below. First we would assume that any preventive maintenance would have been completed while the spare was not needed. When a failure occurs in the stressed component, the spare is plugged in to take over. However the rerun time associated with the failure would still have to be spent. Preventive maintenance on the normally-stressed component could then be done while the spare is active:



.60 .65 .70 .75 .80 .85 .90 .95 1.00

Therefore we see that the spare is normally called upon only during the prevention and down time of the normally-stressed component. However the spare has its own proportion of down and rerun time which will occur during the maintenance time of the normally-stressed component.

Thus we can define a quantity called Spare Need which is the proportion of the normally-stressed component stress time that the spare, excluding spare prevention, is called upon to operate. Continuous Availability, Delay/Stress, and Maintenance/Stress can then be calculated in terms of this Spare Need as indicated. It is well to keep in mind that this is a mean spare need, and may be less than the desirable spare capacity.

Continued in right-hand text.

Spare Need $R_n = (u_p + u_d)/(1 - u_p)$ = .1313

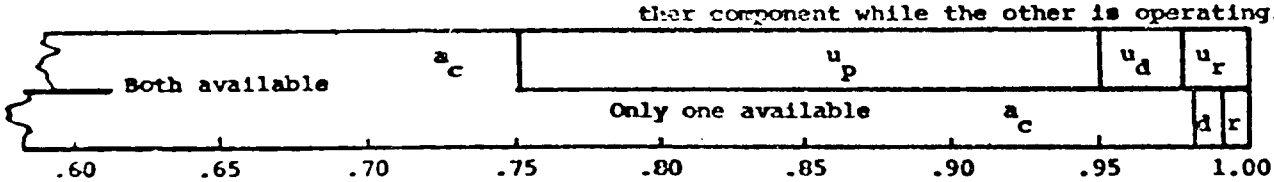
Continuous Availability
$a_{cg} = a_c(1 + R_n)$ = .9800

Delay/Stress
$u_{yg} = 1 - a_c(1 + R_n)$ = .0200

Maintenance/Stress
$u_{mg} = (u_p + u_d)(1 + R_n)$ = .1337

Note that in the example Continuous Availability has been raised from 87 to 98%, Delay reduced from 3 to 2%, and Maintenance increased from 12 to 13%.

Two Stressed Equal Components: We may use the binomial expansion terms to obtain the equations and example figures shown below. The Availability bar-graph shows that both components will probably be operating up to 75% of the time, and only one will probably be available from 75 to 98% of the time. Preventive maintenence would be done on either component while the other is operating.



.60 .65 .70 .75 .80 .85 .90 .95 1.00

Continuous Availability of both components:

$a_{cg2} = a_c^2$ = .7565

Of one or both $a_{cg} = 2a_c - a_c^2$ = .9821

Delay/Stress $u_{yg} = 1 - a_{cg}$ = .0179

Maint./Stress $u_{mg} = 2(u_p + u_d)$ = .2364

In this case Continuous Availability is only 0.2% better than with a spare. Delay is about 10% lower (better) than with a spare, but maintenance is almost twice that with a spare.

However in some circumstances there can be a very important advantage to the use of two stressed equal components. If the system can be designed to utilize the outputs of either or both components, perhaps on different input channels, and perhaps providing different acceptable modes or levels of system output, the result can be better than with a spare. The Availability is higher, the Delay and Maintenance per unit of output are lower, and the component investment is lower per unit of output.

MAJOR COMPONENT INDICES

As an example of the kind of information that can be made available to the system designer, there follows a consolidated analysis of reliability data taken from the operators log on a pair of data-processing systems covering a ten-month working period. It must be emphasized that this information comes from just two specific system installations and therefore cannot be considered as a mean for the 100-odd installations of identical systems. Such reliability information is dependent upon the way the user operates his system as well as upon its inherent reliability. Continued at the left below.

Component	Failures	Range, 90% Confidence		Stress Life	Prevention /Stress	Down/ Stress	Rerun/ Stress	Continuous Availability
	f	- %	+	T_z hr	u_p %	u_d %	u_r %	a_c %
Central Processor	119	13	17	83.4	3.799	.810	.472	94.92
Drum	19	29	53	227.2	1.165	.259	.181	98.40
Tape Control	12	34	73	826.9	.489	.038	.102	99.37
Tape Handler	188	11	13	325.3	.998	.184	.155	98.66
Tape	192	11	13	318.8	0	0	.302	99.70
Reader	45	20	31	97.8	2.065	.545	.170	97.22
Punch	3	52	hi	1397.	.396	.098	.019	99.49
Printer	8	39	100	543.5	.281	.021	.058	99.64
HARDWARE (Model)	694			17.0	10.105	2.260	3.306	83.54
Programmer	424	7	9	20.4	0	.013	.696	99.29
Operator	59	18	26	251.4	.217	0	.399	99.38
Ambient (air cond.)	11	35	78	1309.	.111	.083	0	99.81
Utilities (power)	4	48	190	2480	.081	.020	0	99.90
Inputs (data)	169	11	14	43.9	0	1.389	.224	98.39
SITUATION (Model)	2846			8.2	.548	1.241	4.149	93.84
SYSTEM OPERATION	3540			5.5	10.653	3.501	7.455	78.39

Only the independent failures causing system failure are recorded. The number f of them constitutes a direct indication of the confidence we may have in Stress Life estimates and in other indices derived from Stress Life.

Down percent of stress time is the percent of component stress time that the system was down due to failure of that component. Rerun percent of stress time is the percent of component stress time that the system was occupied in rerun due to failure in that component. Continuous Availability is 100% less Prevention, Down, and Rerun.

Situation indices are treated identically with hardware components to show feasibility of the technique. However we are far less confident of the figures, because the data was not taken with this application in mind.

With clear reliability index objectives in mind, it is not difficult to set up a relatively foolproof and simple logging system, so that we may have adequate confidence in the data collected. Such a system will be suggested in a future paper.

SYSTEM RELIABILITY MODEL

As system failures occur, one must obtain precise component failure and maintenance information so the designer can correct deficiencies and improve the next system design. But he can only do so if he has a workable means of converting component reliability information into system reliability prediction. Reliability models have come into use to solve this problem for complex systems.

This section presents a generalized reliability model which may be used manually or computer-programmed. The actual data resulting from ten months operation (well-seasoned by several prior years of operation) of two data-processing systems is used in a step-by-step example.

RELIABILITY MODEL PROGRAM: Most systems contain quantities of identical components as well as those of which only one is used. The quantity Unavailabilities are somewhat less than individual component Unavailability-times-quantity, because of the overlap effect.

Many systems also contain components stressed only a fraction of the time that the system as a whole is stressed. Thus the inherent Unavailability of such components must be multiplied by this fraction to obtain the contribution to system Unavailability. Let

u_p = Component Prevention/Stress time

u_d = " Down/Stress time

u_r = " Rerun/Stress time

q = Quantity of identical components

R_s = Component/system stress time ratio

t = Task time for complete system

T_f = Component Stress Life, if applicable

1. Group Unavailabilities are computed first, using formulas which properly account for component quantity and Stress Ratio. Prevention/Stress, Down/Stress, and Rerun/Stress times, so far as system Unavailability is concerned, are computed using the formulas shown. These formulas simply convert component Unavailabilities to Availabilities, use the product rule to combine them according to quantity, and then convert back to group Unavailability

$$a_d = 1 - u_d - u_r$$

$$a_c = 1 - u_p - u_d - u_r$$

Prevention/Stress

$$u_{pg} = R_s(a_d^q - a_c^q)$$

Down/Stress

$$u_{dg} = u_d R_s(1 + a_d^q)/(1 - a_d)$$

Rerun/Stress

$$u_{rg} = u_r R_s(1 + a_d^q)/(1 - a_d)$$

However we are also interested in the actual maintenance time required, which the overlap effect does not reduce, so Prevention and Down Maintenance Unavailabilities are computed separately:

Preventive maintenance $m_p = u_p q R_s$

Down maintenance $m_d = u_d q R_s$

2. Group Availabilities are computed then, simply by subtracting Unavailabilities from unity. Group Survival contributions are computed on the basis of their measured Stress Life during the proportion of time they are stressed, and infinite Stress Life when they are not. This results in an upward adjustment of the probability of no failure.

Demand Availability

$$a_{dg} = 1 - u_{dg} - u_{rg}$$

Continuous Availability

$$a_{cg} = 1 - u_{pg} - u_{dg} - u_{rg}$$

Survival $a_{tg} = 1 - R_s(1 - e^{-qt/T_f})$

3. System Availability is then computed using the product rule to combine all group Availabilities. Demand Availability is that assuming preventive maintenance may be interrupted upon demand. Continuous Availability is the long-term Availability taking prevention, down, and rerun time into account. System Survival is the probability of no failure during the task, assuming a successful task start, taking into account the part-time stress of some components.

Demand Availability A_d = product all a_{dg}

Continuous Availability A_c = product all a_{cg}

Survival A_t = product all a_{tg}

$$R_y = (1 - A_d)/(\Sigma u_{dg} + \Sigma u_{rg})$$

Prevention/Stress $U_p = A_d - A_c$

Down/Stress $U_d = R_y \Sigma u_{dg}$

Rerun/Stress $U_r = R_y \Sigma u_{rg}$

Prevention/Stress is of course the difference between Demand and Continuous Availabilities. This total system figure, taken as a ratio to the summation of group Prevention/Stress, can be used to convert individual group Prevention/Stress figures to individual group contributions to system Prevention/Stress. This is done in the table below.

Similarly the system sum of Down and Rerun/Stress time, or $1 - A_d$, may be taken as a ratio R_y to the summation of individual group Down and Rerun/Stress. This ratio may be used to convert individual group Down and Rerun/Stress to their individual group contributions to system Down and Rerun/Stress. The summation of these contributions in the table are the system Down and Rerun/Stress times.

SYSTEM RELIABILITY TABLE: Using the component reliability indices averaging both systems, together with stress time information from the operators logs, the following (Continued to the right) values have been computed with the system reliability model. Note that the system operated 16.5 hours per day, but individual components were stressed for varying proportions of this time.

Qty q	Component	Stress Time T_d hr	Stress Ratio R_s	System Maintenance Prevnt m_p %	System Maintenance Down m_d %	Survival 1 hr A_t	Availability Demand a_{dg} %	Availability Contin a_{cg} %	System Prevn/ Stress U_p %	System Down/ Stress U_d %	System Rerun/ Stress U_r %
1	Cent Proc	16.5	1.000	3.799	.810	98.80	98.718	94.919	3.288	.773	.450
1	Drum	7.2	.436	.508	.113	99.81	99.808	99.300	.440	.108	.075
1	Tape Cont	16.5	1.000	.489	.038	99.88	99.860	99.371	.423	.036	.097
10	Tape Hand	10.1	.612	6.108	1.126	98.15	97.951	92.287	4.902	1.061	.994
10	Tapes	10.1	.612	0	0	98.11	98.307	98.307	0	0	1.616
1	Reader	7.4	.448	.925	.244	99.54	99.680	96.755	.801	.233	.073
1	Punch	7.0	.424	.168	.042	99.97	99.877	99.709	.145	.040	.077
1	Printer	7.2	.436	.123	.009	99.92	99.966	99.843	.106	.009	.024
HARDWARE SYSTM		16.5	1.000	12.120	2.382	94.30	94.292	83.541	10.105	2.260	3.306
6	Programmr	14.4	.873	0	.068	95.83	96.352	96.352	0	.064	3.417
2	Operators	12.4	.752	.326	0	99.70	99.401	99.075	.282	0	.572
1	Ambient	33.6	2.036	.226	.169	99.92	99.831	99.605	.196	.161	0
1	Utilities	16.5	1.000	.081	.020	99.96	99.980	99.899	.070	.019	0
1	Input	12.4	.752	0	1.045	98.31	98.787	98.787	0	.997	.160
SITUATION SYST		16.5	1.000	.633	1.302	88.47	94.434	93.835	.548	1.241	4.149
SYSTEM OPERATN		16.5	1.000	12.753	3.684	83.43	89.044 $=A_d$	78.391 $=A_c$	10.653	3.501	7.455
				Maintenance U_m = 16.437%					System Delay U_y = 10.956%		
				Component addition totals =			88.270	75.517	12.753	3.684	8.046

Individual component maintenance figures, multiplied by their quantity and Stress Ratio, appear in the System Maintenance columns. These add directly to a total maintenance time of 16.4% of the system stress time. Thus total maintenance amounts to 2.7 hours of the 16.5-hour day.

Hardware Availabilities are multiplied to obtain the hardware system Availability, and the same with situation Availabilities. The product of these two is then Availability of the operating system as a whole, computed for both Demand and Continuous consideration.

System Demand Availability is 89%, the Delay being split about equally between hardware and situation. System Continuous Availability is 78%, with hardware Unavailability the principal contributor.

Survival for a one-hour task time, A_t, indicates that the probability of failure is twice as great in the "situation" as it is in the hardware itself.

Component group contributions to system Unreliability are shown in the last three columns, their totals constituting the Indices for this system as a whole. System Delay is the sum of Down and Rerun/Stress, or 11% of the system stress time. This is 1.8 hours of the 16.5-hour day.

Component addition totals are shown at the bottom, exhibiting the pessimistic error with respect to the actual "System Operation" figures.

This paper is designed to lay the reliability groundwork for some economy models. A subsequent paper (reference 1) employs the Availability Indices developed herein, together with performance indices and users cost elements, to present a generalized but practical system economy model. Published data processing component lease costs are used in the complete system example to demonstrate the effectiveness of the technique.

Another paper (reference 2) outlines a regular reporting system, utilizing the model output, to provide current reports of progress (or lack of it) to project management.

References: (1) E. S. Winlund, "Economy models for System Design Engineers" Dec. 15, 1959, (2) E. S. Winlund, "Product Assurance" Jan. 5, 1960, presented at IRE 1960 Winter Convention on Military Electronics. Both are available from the author.

PERFORMANCE OF A ONE UNIT SYSTEM

By: Richard E. Barlow and Larry C. Hunter
Electronic Defense Laboratories, Sylvania Electric Products, Inc.

ABSTRACT

This paper considers an electronic system which upon failure is repaired. Only two states -- the "on" state and the "off" state -- are distinguished. The time-to-failure and the time-to-repair are random variables with general distribution functions. The purpose of the paper is to make available, in a single unified treatment, all mathematical information which is relevant for a reliability analysis of a one-unit system. The following questions are answered:

(a) What is the probability that the system will be on at any given time t?
(b) What is the probability that it will be on t hours or more during a given time interval (0,T)?
(c) What is the expected fractional amount of time that it will be on during (0,T)?
(d) What is the probability distribution of the number of failures during a given time interval?
(e) What is the expected number of failures during a given time interval?

In particular, it is noted that "down" time is approximately normally distributed for large time intervals. The special cases of exponential failure with exponential and constant repair are given as examples.

I. INTRODUCTION

The purpose of this paper is to make available, in a single unified treatment, all of the mathematical information which is relevant for a reliability analysis of a one-unit system. In this paper the word "one unit" shall mean that the system is viewed collectively and that only the performance of the system as a whole is considered -- not the performance of the individual components. The system itself can be composed of one device or many dissimilar devices. A single vacuum tube is a one-unit system. A complex receiver, or even an entire surveillance system can be viewed as a one-unit system when the failure of any component is interpreted as failure of the entire system. Hence we will assume that if a failure occurs the system is completely inoperative, i.e., is in the "failed" state. Repair commences immediately upon failure and, once repaired, the system is returned to the "operating" state. The distribution functions for the time-to-failure and the time-for-repair will be assumed known. The most common assumption is that of an exponential failure distribution and an exponential or constant repair distribution.

The problem concerning the one-unit system is to determine (1) The distribution of the number of failures in (0,t) and its moments, (2) The distribution of the number of completed repairs in (0,t) and its moments, and (3) The distribution of down time in (0,t) and its moments. The probability that the system is operating at any specified time can be computed in terms of the means of the first two distributions.

Also of interest is the probability that the system is operating at a specified time and will continue to operate for a specified interval of time. This quantity, which we call Interval Reliability, is especially important for equipment which must be working when an emergency situation arises.

The stochastic process which we study is known in probability literature as a two-state semi-Markov process. Double generating functions for the distributions listed as (1) and (2) have been obtained by Pyke.[1] The exponential case for the one-unit system has been treated by numerous authors.[2,3] Weiss[4] considered the general problem in connection with the coincidence of periodic pulse trains coming from radar systems. Takács[5] treated the case of exponential failure and general repair distributions as a Type I counter problem. The sojourn time problem which we referred to as (3) has been exhaustively studied by Takács.[6,7,8]

The new feature of this paper is that it unifies previous related work and identifies it with the reliability problem. We present, as examples, the special cases of exponential failure with exponential and constant repair.

II. THE DISTRIBUTION OF THE NUMBER OF FAILURES

We will need some notation. Let F always denote the failure distribution and G the repair distribution. Let H denote their convolution, i.e.

$$H(t) = \int_0^t G(t-x)\, dF(x).$$

It will be convenient to label the operating state by 0 and the failed state by 1. Let $N_{ij}(t)$ denote the number of visits to state j in (0,t) given that the system enters state i at time 0. Then $N_{01}(t)$ is the number of failures in (0,t) if the system is put into operation at time t=0. Let X with or without subscripts be a random variable denoting the time-to-failure and Y be a random variable denoting the time-for-repair. We obtain

$$P[N_{01}(t) = k] = P[X+Y_1+X_1+\ldots+Y_{k-1}+X_{k-1} \le t] - P[X+Y_1+X_1+\ldots+Y_k+X_k \le t]$$

$$= F*H^{(k-1)}(t) - F*H^{(k)}(t) \qquad (k \ge 1)$$

and

$$P[N_{01}(t) = 0] = 1 - F(t)$$

where * denotes pairwise convolution and $H^{(k)}(t)$ means that H is convoluted with itself k times. We can obtain a double generating function for the distribution of $N_{01}(t)$ as follows:

Let

$$\phi_{01}(z,t) = \sum_{k=0}^{\infty} z^k P[N_{01}(t) = k]$$

and

$$\Phi_{01}(z,s) = \int_0^{\infty} e^{-st} d_t \phi_{01}(z,t).$$

Then

$$\Phi_{01}(z,s) = \frac{zF^*(s)(1-z)}{[1 - zH^*(s)]}$$

where upper * denotes Laplace-Stieltjes transform. Let

$$\Psi_{ij}(t) = E[N_{ij}(t)],$$

and

$$\Psi^*_{ij}(s) = \int_0^{\infty} e^{-st} d\Psi_{ij}(t);$$

then,

$$\Psi^*_{01}(s) = \frac{d}{dz} \Phi_{01}(z,s)\Big|_{z=1},$$

so that

$$\Psi^*_{01}(s) = \frac{F^*(s)}{1-F^*(s)\,G^*(s)} \qquad (2.1)$$

In a similar manner we obtain

$$\Phi_{00}(z,s) = \frac{1-H^*(s)}{1-zH^*(s)}$$

and

$$\Psi^*_{00}(s) = \frac{F^*(s)\,G^*(s)}{1-F^*(s)\,G^*(s)} \qquad (2.2)$$

Equations (2.1) and (2.2) give the Laplace-Stieltjes transform of the expected number of failures and repairs, respectively, in time t. Since it is often difficult to invert these expressions, it is worthwhile to also derive integral equations for these expressions. $\Psi_{ij}(t)$ is the expected number of visits to state j in time t if at time t=0 the system enters state i. Note that if the system is on at time 0, then the expected number of visits to the "on" state, given that first failure occurs at time x, is $\Psi_{10}(t-x)$. Therefore

$$\Psi_{00}(t) = \int_0^t \Psi_{10}(t-x)\, dF(x). \qquad (2.3)$$

Similarly, if the system is off at time 0, the expected number of visits to the "on" state, given that the first failure occurs at time x, is $1+\Psi_{00}(t-x)$. Hence

$$\Psi_{10}(t) = \int_0^t [1 + \Psi_{00}(t-x)]\, dG(x). \qquad (2.4)$$

Knowing F(x) and G(x), equations (2.3) and (2.4) can be used to determine $\Psi_{00}(t)$ and $\Psi_{10}(t)$. Note that by taking Laplace-Stieltjes transforms, we obtain (2.1).

In a similar manner we have

$$\Psi_{01}(t) = \int_0^t [1 + \Psi_{11}(t-x)]\, dF(x) \qquad (2.5)$$

$$\Psi_{11}(t) = \int_0^t \Psi_{01}(t-x)\, dG(x) \qquad (2.6)$$

which can be solved for $\Psi_{01}(t)$ and $\Psi_{11}(t)$ numerically.

Let $P_{ij}(t)$ denote the probability that the system is in state j at time t if it starts in state i at t=0. Note that

$$N_{01}(t) - N_{00}(t) = \begin{cases} 1 \text{ if system is off at time } t. \\ 0 \text{ otherwise.} \end{cases}$$

Hence

$$P_{01}(t) = E[N_{01}(t) - N_{00}(t)]$$

$$= \Psi_{01}(t) - \Psi_{00}(t) \qquad (2.7)$$

and

$$P_{00}(t) = 1 - P_{01}(t).$$

Similarly

$$P_{10}(t) = \Psi_{10}(t) - \Psi_{11}(t), \qquad (2.8)$$

and

$$P_{11}(t) = 1 - P_{10}(t).$$

Since $N_{01}(t)-N_{00}(t)$ is a random variable which assumes only the values 0 and 1, we note that

$$[N_{01}(t)-N_{00}(t)]^n = N_{01}(t)-N_{00}(t)$$

where n is any positive integer. In particular, if n=2,

$$E[N_{01}(t)-N_{00}(t)]^2 = \Psi_{01}(t)-\Psi_{00}(t)$$

and

$$2E[N_{01}(t)\, N_{00}(t)] = E[N_{01}(t)^2] + E[N_{00}(t)^2] + \Psi_{00}(t) - \Psi_{01}(t).$$

In this manner we can compute the correlation coefficient of $N_{01}(t)$ and $N_{00}(t)$ in terms of their respective first and second moments.

The Laplace transform of the distribution of the number of failures can be obtained in the following way. Let $W(t,n)$ be the probability of n or less failures in time t starting with the unit on. Then

$$W(t,n) = \sum_{k=0}^{n} P[N_{01}(t) = k]$$

$$= 1-F(t) + \sum_{k=1}^{n} [F*H^{(k-1)}(t) - F*H^{(k)}(t)]. \qquad (2.9)$$

Then

$$W^*(s,n) = \int_0^\infty e^{-st}\, d_t\, W(t,n)$$

$$= -F^*(s)\,[H^*(s)]^n. \qquad (2.10)$$

In general these Laplace-Stieltjes transforms are difficult to invert. However, some asymptotic formulae are available from renewal theory.[9] In particular

$$\Psi_{01}(t) = \frac{t}{\mu} - \frac{\xi}{\mu} + \frac{\mu^{(2)}}{2\mu^2} + o(1) \qquad (2.11)$$

$$\Psi_{00}(t) = \frac{t}{\mu} - 1 + \frac{\mu^{(2)}}{2\mu^2} + o(1) \qquad (2.12)$$

and

$$P_{01}(t) = \frac{\eta}{\mu} + o(1) \qquad (2.13)$$

where ξ and η are the means of F and G respectively and $\mu = \xi + \eta$ and $\mu^{(2)}$ are the first and second moments of H.

SUMMARY OF SECTION II

1. The Laplace-Stieltjes transforms of the expected number of failures and repairs are given by equations (2.1) and (2.2), respectively.

2. Integral equations for the expected number of failures and repairs are given in equations (2.3), (2.4), (2.5), (2.6).

3. The probability the system is on at time t is given by (2.7) and (2.8)

4. The probability of n or less failures in time t is given by (2.9) and its transform by (2.10).

5. Asymptotic expressions of these qualities are given in (2.11), (2.12), and (2.13).

Example 2.1

Suppose the failure distribution is exponential, i.e., $F(t) = 1 - e^{-at}$ and the time-for-repair is also exponential, i.e., $G(t) = 1 - e^{-bt}$. Then

$$F^*(s) = \frac{a}{s+a}$$

and

$$G^*(s) = \frac{b}{s+b}.$$

The Laplace-Stieltjes transform of the expected number of failures is, by (2.1)

$$\Psi_{01}^*(s) = \frac{a(s+b)}{s^2+(a+b)s}$$

Consulting a table of Laplace transforms we see that

$$\Psi_{01}'(t) = ae^{-\frac{a+b}{2}t}\left\{\cosh\frac{(a+b)}{2}t + \frac{b-a}{a+b}\sinh\frac{a+b}{2}t\right\}$$

where $\Psi_{01}'(t)$ denotes the derivative of $\Psi_{01}(t)$. Hence

$$\Psi_{01}(t) = \frac{abt}{a+b} - \frac{ae^{-(a+b)t}}{2(a+b)} + \frac{a(b-a)}{2(a+b)^2}\, e^{-(a+b)t} + c$$

where c is a constant of integration. Since

$$\Psi_{01}(0) = 0, \quad c = \frac{a^2}{(a+b)^2}$$

and

$$\Psi_{01}(t) = \frac{a^2}{(a+b)^2} + \frac{abt}{a+b} - \frac{a^2}{(a+b)^2} e^{-(a+b)t}$$

In a similar manner, we compute

$$\Psi_{00}(t) = \frac{-ab}{(a+b)^2} + \frac{abt}{a+b} + \frac{ab}{(a+b)^2} e^{-(a+b)t}.$$

Thus

$$P_{01}(t) = \Psi_{01}(t) - \Psi_{00}(t)$$

$$= \frac{a}{a+b} - \frac{a}{a+b} e^{-(a+b)t}$$

and

$$P_{00}(t) = \frac{b}{a+b} + \frac{a}{a+b} e^{-(a+b)t}.$$

These probabilities could also have been obtained in a direct fashion. Upon inversion of (2.10), we obtain

$$W(t,n) = \sum_{j=1}^{n+1} \frac{A_j t^{j-1}}{(j-1)!} e^{-at} + \sum_{j=1}^{n} \frac{B_j t^{j-1} e^{-bt}}{(j-1)!}$$

where

$$A_j = a^{j-1} + \sum_{k=1}^{n} (-1)^{n+1-j} \binom{n+k-j}{k-1} \frac{b^{k-1} a^{n+1}}{(b-a)^{n+k-j+1}}$$

$$B_j = b^{j-1} + \sum_{k=1}^{n+1} (-1)^k \binom{n+k-j+1}{k-1} \frac{a^{k-1} b^{n+2k-2j}}{(b-a)^{n+k-j}}.$$

This quantity was obtained as Example 2 in ref. 5.

Example 2.2

Suppose again that the failure distribution is exponential, but that the time-for-repair is constant, say β. Then

$$F^*(s) = \frac{a}{a+s}$$

and

$$G^*(s) = e^{-\beta s}.$$

The mean number of failures in $(0,t)$ becomes

$$\Psi_{01}(t) = \left[\frac{t}{\beta}\right] + 1 + \sum_{j=0}^{\left[\frac{t}{\beta}\right]} \sum_{k=0}^{j} a^k \frac{(t-\beta j)^k}{k!} e^{-a(t-\beta j)},$$

and

$$\Psi_{00}(t) = \left[\frac{t}{\beta}\right] + \sum_{j=0}^{\left[\frac{t}{\beta}\right]} \sum_{k=0}^{j} a^k \frac{(t-\beta j)^k}{k!} e^{-a(t-\beta j)} + \sum_{k=0}^{\left[\frac{t}{\beta}\right]} e^{-a(t-\beta k)} \frac{a^k}{k!} (t-\beta k)^k$$

where $\left[\frac{t}{\beta}\right]$ denotes greatest integer contained in t/β. Hence

$$P_{01}(t) = 1 - \sum_{j=0}^{\left[\frac{t}{\beta}\right]} a^j \frac{(t-\beta j)^j}{j!} e^{-a(t-\beta j)}$$

and

$$P_{00}(t) = 1 - P_{01}(t).$$

The distribution of the number of failures can be obtained as

$$W(t,n) = \begin{cases} \sum_{j=1}^{n+1} a^{j-1} \frac{(t-n\beta)^{j-1}}{(j-1)!} e^{-a(t-n\beta)} & \text{if } n\beta \le t \\ 1 & \text{if } n\beta > t \end{cases}$$

These quantities were derived in reference 5.

III. DISTRIBUTION OF SYSTEM DOWN TIME

Let $\beta(t)$ denote the amount of time the system is down during $[0,t]$. That is, $\beta(t)$ is the amount of time spent in repairing the system in time t. Put

$$\Omega(t,x) = \Pr[\beta(t) \le x].$$

L. Takács showed that the distribution function for $\beta(t)$ is:

$$\Omega(t,x) = \sum_{n=0}^{\infty} G^n(x)\left[F^n(t-x) - F^{n+1}(t-x)\right] \qquad (3.1)$$

where $G^n(x)$ and $F^n(x)$ denote the n-times iterated convolution of the distribution function $G(x)$ and $F(x)$, respectively, with itself ($G^0(x) = 1$ if $x \ge 0$, $G^0(x) = 0$ if $x < 0$ and $F^0(x) = 1$). Further, he proves the following,

THEOREM 1

If the variance of X and Y, σ_α^2 and σ_β^2 respectively, is finite, then putting $E(X) = \alpha$ and $E(Y) = \beta$, we have

$$\lim_{t\to\infty} \Pr\left\{ \frac{\beta(t) - \frac{\beta t}{\alpha+\beta}}{\sqrt{\frac{\beta^2\sigma_\alpha^2 + \alpha^2\sigma_\beta^2}{(\alpha+\beta)^3}\beta\, t}} \le x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp -\frac{u^2}{2}\, du.$$

Example 3.1

Suppose we wish to compute the probability that the system is down more than 24 hours in 10,000 hours of operation. The failure and repair distributions are unknown, but from sample data, the following estimates of means and variances are made:

$$\alpha = 1{,}000 \text{ hours} \qquad \sigma_\alpha^2 = 100{,}000$$

$$\beta = 2 \text{ hours} \qquad \sigma_\beta^2 = 4 .$$

Appealing to Theorem 1, we know that for t=10,000

$$\frac{\beta(t) - \frac{\beta t}{\alpha+\beta}}{\sqrt{\frac{\beta^2\sigma_\alpha^2 + \alpha^2\sigma_\beta^2}{(\alpha+\beta)^3}\beta\, t}} = \frac{\beta(10{,}000) - 19.96}{6.66}$$

is approximately normally distributed with mean 0 and variance 1. Hence

$$\Pr\{\beta(10{,}000) \ge 24\} = \Pr\left\{\frac{\beta - 19.96}{6.66} \ge \frac{24 - 19.96}{6.66}\right\} \approx \frac{1}{\sqrt{2\pi}} \int_{.6}^{\infty} \exp\frac{-u}{2}\, du$$

$$\approx .28 .$$

Thus, the probability the system is down more than 24 hours in 10,000 hours is about .28. Note that the probability that down time will exceed 30 hours is less than .07.

$$\Pr\{\beta(10{,}000) \ge 30\} < .07.$$

Example 3.2

It is easy to work out the exact distribution for the down time, given by equation (3.1), when Y(t) and G(t) are both exponentials:[6]

$$F(t) = 1 - e^{-at}, \qquad G(t) = 1 - e^{-bt} .$$

Then

$$F^{(n)}(t-x) - F^{(n+1)}(t-x) = e^{-a(t-x)}\, \frac{[a(t-x)]^n}{n!}$$

and

$$Q(u+x,x) = \sum_{n=0}^{\infty} e^{-au}\, \frac{(au)^n}{n!}\, G^{(n)}(x) .$$

The Laplace-Stieltjes transform is

$$\int_0^\infty e^{-sx}\, d_x\, Q(u+x,x) = e^{-au[1-R(s)]}$$

where

$$R(s) = \int_0^\infty e^{-sx}\, dG(x) = \frac{b}{b+s} .$$

So by inversion

$$Q(u+x,x) = e^{-au}\left[1 + \sqrt{abu}\int_0^x e^{-by} y^{-\frac{1}{2}} I_1(2\sqrt{abuy})\, dy\right]$$

where $I_1(x)$ is the Bessel function of order 1 for the imaginary argument defined by

$$I_1(x) = \sum_{j=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{2j+1}}{j!(j+1)!} .$$

By substituting u = t-x we obtain

$$Q(t,x) = e^{-a(t-x)}\left[1 + \sqrt{ab(t-x)} \cdot \int_0^x e^{-by}\, y^{-(1/2)}\, I_1(2\sqrt{ab(t-x)y})\, dy\right].$$

IV. INTERVAL RELIABILITY

As a final result we obtain the Interval Reliability of a one-unit system. Interval reliability is the probability that at a specified time, the system is operating and will continue to operate for an interval of duration, say x. The system may, for example, be a safety device which is used only when an emergency situation arises. It may, however, still fail while not being used. This is particularly apropos to surveillance equipment. We may then ask for the probability that a system will survive for an interval of length x without repair when called into use at time t. Let R(x,t) denote this probability. Then

$$R(x,t) = 1 - F(t+x) + \int_0^t [1 - F(t-y+x)]\, d\Psi_{00}(y)$$

since $d\Psi_{00}(y)$ is the probability of a completed repair at time y. The limit as t approaches infinity of this quantity follows from the key renewal theorem,[9] i.e.,

$$\lim_{t\to\infty} R(x,t) = \frac{\int_x^\infty [1 - F(y)]\,dy}{\xi + \eta}.$$

We may also ask for the probability that the system will be repaired within a sufficiently short interval if it is not operating at a specified time. We will call this quantity the servicing reliability and use the notation $S(x,t)$. Thus $S(x,t)$ is the probability that either the system is operating at time t or, if it is not operating, that it will be repaired within an interval of length x. Reasoning as above we obtain:

$$S(x,t) = 1 - \int_0^t [1 - G(t-y-x)]\,d\Phi_{01}(y).$$

Again applying the key renewal theorem, we obtain

$$\lim_{t\to\infty} S(x,t) = \frac{\int_x^\infty [1 - G(y)]\,dy}{\xi + \eta}.$$

Example 4.1

Again consider the exponential case, where $F(t) = 1-e^{-at}$, and $G(x) = 1-e^{-bt}$. Then

$$R(x,t) = \frac{a}{a+b} e^{-a(t+x)} + \frac{be^{-ax}}{a+b}$$

$$+ \frac{a^2}{b(a+b)} (1-e^{-bt})\, e^{-(t+x)}.$$

Clearly

$$\lim_{t\to\infty} R(x,t) = \frac{be^{-ax}}{a+b}$$

which one would expect. Also,

$$S(x,t) = 1 - \frac{a}{a+b} e^{-bx} - \frac{a}{a+b} e^{-(a+b)t}\, e^{-bx}$$

and

$$\lim_{t\to\infty} S(x,t) = 1 - \frac{ae^{-bx}}{a+b}.$$

REFERENCES

1. Pyke, R., "Markov Renewal Processes with Finitely Many States," Columbia University Technical Report, ONR Contract Number Nonr-266(59).

2. Feller, W., "An Introduction to Probability Theory and Its Applications," Wiley, 1957.

3. Hosford, J., "Measures of Dependability," Operations Research, Vol 8, Number 1, Jan. 1960.

4. Weiss, G., "On the Coincidence of Certain Random Functions," Quart. Appl. Math., XIV-1, 103, (1956).

5. Takács, L., "Occurrence and Coincidence Phenomena in Case of Happenings with Arbitrary Distribution Law of Duration," Acta Math. Acad. Sci. Hung. 2, pp. 275-298, 1951.

6. Takács, L., "On Certain Sojourn Time Problems in the Theory of Stochastic Processes," Acta Math. Acad. Sci. Hung., pp. 169-191, 1957.

7. Takács, L., "On Limiting Distributions Concerning a Sojourn Time Problem," Acta Math. Acad. Sci. Hung., pp. 279-294, 1957.

8. Takács, L., "On a Sojourn Time Problem in the Theory of Stochastic Processes," Trans. Amer. Math. Soc., Vol 93, No. 3, December 1959.

9. Smith, W., "Asymptotic Renewal Theorems," Proceedings of the Royal Society of Edinburgh, Vol XXIV, pp. 9-48, 1954.

STATISTICAL PITFALLS FOR THE RELIABILITY ENGINEER

By: G. H. Beckhart, Product Assurance, Radio Corporation of America

Introduction: The field of Reliability has developed phenomenally during the past few years. Some aspects of development have been extremely fruitful, other aspects have not returned the dividends expected from the effort expended. The lack of return is due, in part, to semantic difficulties between engineers and mathematicians. This presentation will concern itself with four areas of the semantic hurdles that face the reliability engineer.

1: The word "efficiency" is one of the key factors in engineer/statistician relations, as this word is used by both groups with entirely different meanings. Statisticians spend a great deal of time talking about the statistical efficiency of a particular technique, and engineers assume, usually erroneously, that the statistician is referring to engineering efficiency.

The statistician calls a statistic efficient if it has minimum variance. Efficiency is a relative term meaning that a particular statistic has less variability than any others when computed from samples drawn at random from a bowl of random numbers or chips numbered to reproduce a specific density function. Thus, the efficiency may be a function of the density functions.

The theoretical advantage of the so-called efficient statistic is that a smaller sample is required to estimate the parameters of the distribution than would be required by the use of any other statistic.

The desirability of using efficient statistics is reiterated so often in statistical text books that non-statisticians have joined the chorus. Hence, in books written by engineers, you will find words to the wise concerning the desirability of using the standard deviation, due to its efficiency, to estimate variability. The use of the range is, if mentioned, usually scorned.

In practice, it frequently turns out that the standard deviation, while having statistical efficiency, actually is more variable than the range. In experiments conducted by Dixon, at the University of Oregon, college students made so many errors in computing the standard deviation of samples drawn from a bowl of normal numbers, that its variability was actually greater than that of the range.*

The actual desirability of using the range may be judged by the number of statistical research articles on the use of various types of ranges. One range used in life testing is the semi or quasi range which is the time from the first failure to the median failure.

*P. 245 - Dixon and Massey - "Introduction of Statistical Analysis" - McGraw-Hill 1951

The definition of efficiency engineers should employ is that of "Assignable Causes revealed at minimum cost." Hence, the engineer should choose his statistical weapon depending on the assignable cause which he suspects may be present, or which, if present, would be most detrimental to reliability. In particular, this means that statistical techniques chosen to assist in reliability investigations should be those helpful in interpreting the physical mechanism of failure or revealing changes in the physical mechanism. The ability to economically detect assignable causes producing the following effects on data should, in an engineering investigation, be the measure of efficiency - gross errors or blunders, shifts in average, shift of variability, slow fluctuation or trend, fast fluctuation or cycle, and correlations.

Some years ago, statistics of high engineering efficiency were tabulated* with respect to various data influencing effects or assignable causes in a series of articles in Industrial Quality Control. All of these methods are easy to apply and are of great engineering utility in revealing the presence of assignable causes, trends, cycles, wild shots or gross errors, etc. It is generally assumed that if the effects are statistically significant it will be possible to isolate and identify the physical process responsible for influencing the data. Whether this latter assumption is true will depend largely on the skill of the engineer looking for the assignable cause whose presence was revealed by the data.

2: The second point which causes difficulty is in the application of mathematical models. The training of some engineers is so condensed that they do not, in their college days, have sufficient time to examine critically the basic assumption upon which the models are based. Hence, some engineers do not question a statistician too closely as to what his fundamental assumptions are and will inadvertently employ statistical models that are inappropriate for his particular need. In addition, mathematical "one-up-manship" precludes a full discussion by the mathematician of the practical aspects of applying a particular model.

As you are all aware, most equations used in engineering work have to be simplifications of a basic physical model. For example, in designing audio amplifiers, three mathematical models of the amplifiers are employed in analyzing and designing the electronic circuits. Each of these is a simplification of actual circuit and each is useful in a certain frequency range. The use of

*Olmstead, Paul S. - "How to Detect the Type of an Assignable Cause"
Part I - Clues for Particular Types of Trouble Industrial Quality Control - Nov. 1952 Vol. II
Part II - "Procedure when Probable Cause is Unknown" - Industrial Quality Control - January 1953 Vol. II

these models contributes to the design engineer's efficiency. However, a different model must be used if he is going to design a broad band video amplifier, so the engineer must be knowledgeable enough to choose his models with discretion. The most renowned case of the misapplication of a mathematical model in recent years is that of the famous bridge on the West Coast which whipped in the breeze until it fell apart. This was due to an engineer overlooking a simplification in the model that neglected a small effect, which in his application became a large effect.

Reliability Engineers have been depending on the statistician to tell them what model to use. The mathematician, being human, is apt to choose a model which is easy to analyze mathematically, or to choose one because there is no derrogatory information about it. The Reliability Engineer must assure himself that the model chosen really fulfills his particular need. Component failures are generally represented by the exponential distribution. It is reasonably valid, and tables for hypothesis testing are available as well as confidence limits for point estimation. However, the sample sizes and the length of the test required to obtain statistical significance are usually more than may be economically justified. If the exponential represents your failure patterns and you have the money and the time required by the basic assumptions to get the answers - use this model. However, if your aim is to enhance reliability at minimum cost, without necessarily estimating mean-time to failure, use Sobel's models for selecting binomial populations in planning your component tests. This model asks only one question - "Do we have the best population?" It does not ask what is the mean-life of this population. The saving in sample size and time by the use of this model is considerable. Hence, a shift in direction from reliability measurement to reliability improvement may result in increased engineering efficiency for your part selection program. If the equipment is ground equipment amenable to identification of failures and replacement of failed components with little equipment down-time, this may be the way to allocate your funds and manpower for greatest reliability improvement per dollar expended.

Verification of mean-time-to-failure requirement of equipment may cost from three to five times the initial equipment cost when one is attempting verification based on exponential assumptions. Paradoxically the fewer the failures the broader the confidence interval; hence, better equipment tends to require more expensive tests. The economic goal of reliability engineering should be maximum operational effectiveness at minimum cost. This aim should bring the following factors into the reliability model.

Maintenance Cycle - How often must routine maintenance be performed?
Average down-time for routine maintenance.
Average down-time for unscheduled maintenance.
Maximum down-time for unscheduled maintenance.
Maintenance philosophies and skill levels required.
Expected performance degradation and indication of degradation.

Increased emphasis on these factors in your reliability model may reduce the need for estimating mean-time-to-failure. With proper definitions, some of the above factors may be estimated by conventional time study techniques.

The commonly accepted definition of reliability, "The probability of performing without failure a specified function under given conditions for a specified period of time," generated a useful model of reliability for parts or simple black box. However, rigid adherance to this definition leads to models that are not applicable to system performance. One has to define reliability very differently if an adequate systems model is to be devised. An alternative definition* is "Reliability is the probability that a system, function, or mode will provide other than zero performance for a given period of time when subjected to its intended environment."

The beauty of this definition is that models of system performance may be constructed which are amenable to measurement and prediction. Thus, system availability may be measured or system loss may be predicted. Incidentally, system loss is defined as the summation of all possible fractional losses of equipment performance times the respective probability of each of those malfunctions.

Although models for systems have generally ignored repair and have usually assumed component failures are independent; a real system cannot ignore repair,** and sub-systems and parts do not always fail independently. Therefore, a complete physical model of reliability has to reconcile system loss, availability, repair, and dependent failures of parts.

Models of the economics of the total reliability situation have been lacking, or if available, ignored. Possibly this is due to the admitted

* Friddell and Jacks - "System Operational Effectiveness" IRE-ASQC
Proceedings of the 5th National Symposium on Reliability and Quality Control - January 1959

** R. E. Barlow and L. C. Hunter -
System Efficiency and Reliability "Technometrics" - Vol. 2 - #1 - February 1960

* S. S. Gupta, M. J. Hayett, M. Sobels
"Selection and Ranking Problems with Binomial Populations" P. 635
Transactions of 1957 National Convention of the American Society for Quality Control

complexity of the reliability model, on which the economics must be based. For example, operational effectiveness for ground equipment may be achieved at a much lower over-all cost if the product is designed for ease of maintenance and servicing, rather than spending an equivalent sum for improved components, as well as the reliability model. However, this implies knowledge of the maintenance and service costs for the optimum allocation of effort to be achieved.

This leads to a rather involved problem in operations research; for example, how does one measure the economic gain or loss if the system model leads to Sobel's techniques for testing components?

3: The third area that needs improvement is the realization by both the engineer and the statistician that one of the engineer's major problems is making inferences from raw data when there is no statistical model. This problem, is of course, quite different than that of making deductions from a model. Unfortunately, many engineers, as well as mathematicians, do not act as if they are aware of this aspect of the engineer's problem.

This third area is probably the most difficult one of all, and yet is the problem the engineer is continually faced with in his daily activities. He constantly has to make inferences from raw data whose pedigree he doesn't know. Very few discussions on reliability testing or engineering data review the statistical stability of the data. If the data is not statistically stable, or does not come from a process which is statistically stable, you cannot extrapolate present data into prognostications of future performance. The best known test of statistical stability is the control chart devised by Dr. Walter Shewhart for use in analysing engineering data.

The control chart is a graphical procedure based on a combination of sound statistical and engineering principles. Unfortunately, it has been largely ignored by engineers and has usually been applied in analysis of the stability of production data. The use of these charts as an aid in locating and identifying "assignable causes" responsible for poor product performance is well known and extensively documented in this country and abroad.*

Life test data should be tested for stability by the use of control charts.** This will assist in pinpointing test instrumentation problems, changes in rate of degradation of parts, non-homogenity in the sample and other assignable causes. Reliability will only be enhanced if the assignable causes are identified and suitable design or process changes made to prevent their reoccurrence.

In addition to control chart data on life test performance, the maintenance of a consistent reliability level requires that the production process of parts, sub-assemblies, and units be designed and controlled to achieve statistical stability.* Certainly we could not afford to go to these extremes of the Bell Telephone Co. L-3 repeater system. However, system sensitive components certainly should be given this treatment. Otherwise our predictions concerning component behaviour are at best rough guesses.

The rejection or use of wild or outlying observations is one that requires considerable engineering and statistical skill. Are these wild observations part of a stable underlying distribution or part of several underlying distributions? Did we know before the observation that it was likely to be wild,** could we recognise afterwards reasons for its wildness (assignable causes), or could we find no explanation? The engineers' need is, of course, to find an underlying physical explanation responsible for the samples extreme deviation from the norm. The engineer has to determine if there is enough statistical stability to assume a model exists.

4: The fourth area is concerned with the use of prior knowledge. Engineers make decisions concerning component parts on the basis of prior tests, prior field failure information, and any other prior information they can find. The word "prior" is being used in its fullest engineering sense, meaning that all previous knowledge, including statistical knowledge, is being employed. It is extremely desirable that some reasonably rigorous techniques of employing this knowledge be developed. This is particularly necessary in the realm of demonstrating contractual compliance to reliability requirements. At present all we can offer the customer are lengthy, expensive test programs which, if they do demonstrate compliance, may do so after the need for the equipment no longer exists.

To summarize: The four areas requiring improved understanding are concerned with engineering efficiency, mathematical models, absence of models, and the effective use of prior engineering knowledge. Judging by recent symposia, progress is being made in the development of models, but we have a long way to go before statistics with high engineering efficiency are universally employed, or data is tested for stability, or we find out how to utilize prior information.

* Garrett, Tuffnell, Waddell - "The L-3 Coaxial System: Application of Quality Control Requirements in the Manufacture of Components" - Bell System Technical Journal - Vol. XXXII #4 - July 1953

** Enoch Ferrel - P. 700-702 - "Statistical Methods in the Development of Apparatus Life Quality" - Transactions of AIEE - Oct. 1945

* Bell System Technical Journal - Vol. XXXII - #4 July 1953

** William H. Krushal - "Some Remark on Wild Observations" - Technometric Vol. 2 - #1 February 1960

THE USE OF IBM CARDS TO PREDICT, CONTROL, AND MEASURE RELIABILITY OF A MISSILE ELECTRONICS UNIT

By: G. F. Dolan
Engineering Division
Hughes Aircraft Company
Culver City, California

Summary

Electronics unit parts list information is punched on IBM cards at the time the circuit is designed. Typical information which might be included in this design phase listing are: design schematic circuit symbol number, nominal value, component tolerance for satisfactory circuit operation, operating stress, part description, transient stress, failure chart number, lower limit of operating temperature, component failure rate at this temperature, upper limit of operating temperature, failure rate at this temperature, and a circuit identification number.

Circuit changes can be incorporated by adding deleting or modifying cards. Therefore, a current reliability assessment can always be made. Mean-time-between-failures figures can be obtained for individual circuits, types of components or for the entire unit.

The design parts list and reliability information are released to the producitizing section. Here various circuits are combined and packaged on chassis. A transposition list is composed equating design schematic circuit symbol numbers to chassis circuit symbol numbers. This list enables the data processing facility (IBM car facility) to generate a chassis parts list which includes all the original design and reliability information. Mean-time-between-failures figures can then be obtained for each chassis and for the electronics unit. As in the design stage, failure rate tabulations can be obtained for classes of components.

Future plans include (either on decks of IBM cards or on tape) the following libraries: approved parts, approved vendors, and failure rate information at predetermined derating values. Circuit designers will then have to enter on the card punching sheet only the following information: circuit symbol number, nominal value, tolerance, and stress.

It is planned to coordinate this system with a factory and field IBM card failure reporting system.

Introduction

The complexity of today's electronic equipment encourages the use of a machine type data handling program* for the recording and transmission of information connected with the equipment. This paper describes such a program which was instituted primarily to insure the reliability of a missile electronics unit. The paper is divided into three parts. Part one discusses the basic IBM card data handling system for predicting, controlling, and measuring reliability. Part two discusses possible extensions of this basic system. Part three treats in a qualitative manner the criterion to determine whether machine type data handling should be employed.

PART ONE

Predicting and controlling reliability

A diagram of the basic IBM card data handling system is shown in Figure 1.

In the design block diagram phase a rough reliability estimate can be made based on the anticipated number of the various types of components necessary to perform the required functions. At this point there is no application for machine type data handling. When the circuit design is reasonably firm, detailed component reliability information is available. The information should be recorded at this time. If the prototype product design is not done by the circuit design group this information should be transmitted to the product design group. This is the time for choosing a data handling system. It was decided for this project to record component reliability information on IBM cards.

Although the intent was to insure electronics unit reliability, it was recognized that a practical system for doing so must include capability for the other data handling required. Therefore, it was decided to incorporate the component parts list information and reliability information on the same IBM cards. The reliability information would then become an integral part of the design release.

These cards would constitute the "prediction and control deck number 1" shown in Figure 1. Figure 2 shows the sample form that the circuit design engineer would prepare prior to having the deck of cards punched.

* Information can be placed on IBM cards. IBM cards with microfilm inserts, magnetic tape, magnetic cards, etc.

One line on the form is used for each component (schematic circuit symbol number). On this line is included the following information: circuit symbol number, nominal value, tolerance required for satisfactory circuit operation, component predominant operating stress, component description, transient stress, failure rate chart number, derating, lower limit of operating temperature, failure rate at this temperature, upper limit of operating temperature, failure rate at this temperature, and finally the code group designation indicating the circuit in which the component is used. This information can be punched on one IBM card. Figure 3 shows one such card. If additional information is to be recorded, more than one card would be required per component.

If reliability is to be assured, it is essential that the type of information punched on these cards be carried along through the design, production, and field phases of a product. Too often such important reliability information as total component tolerance required during life under all environmental conditions and the transient stress expected, is passed along only implicitly and incompletely. The design engineer recommends a certain type of resistor of capacitor which he feels is suitable. The actual information may be in the mind of the designer or in a file of experimental data. The designers recommendation as to component type may not be followed because of packaging or availability considerations; or the recommended type, used originally, may be changed in some later phase. Therefore, if the original design reliability is to be maintained, this information must be recorded and transmitted during all phases.

It is necessary to include this detailed reliability information in the design release to assure that these aspects of reliability are not neglected. Machine type data handling such as the IBM card system described here make this practical.

At this stage there exists a deck of IBM cards (one for each component) for a particular circuit (code group 03). This deck can generate many types of listings. A page from one such listing, a parts release listing, is shown in Figure 4. Figure 5 to 7 show pages from parts release listings for code groups 01, 02, and 05. These listings include such reliability information as component stresses, recommended derating, and predicted component failure rates. These four code groups or circuits were packaged on a chassis AD*. Other groups were combined to make up other chassis.

A listing can now be obtained from the combined decks of IBM cards for code groups, 01, 02, 03, and 05 which will give the component stresses, recommended derating, and predicted failure rates for all the circuitry to be packaged on chassis AD. The total failure rate for this combined circuitry was 25.13 per cent per thousand hours.

After a chassis schematic had been drawn, a transposition list was made equating development schematic circuit symbols to chassis schematic circuit symbols. By means of this transposition list, the data handling section generated a transposition deck of IBM cards. From this deck and the decks of the four circuit code groups a new deck was punched for chassis AD. All the original component information concerning reliability: tolerance, stress, failure rate, etc. was in this deck. A chassis AD parts listing was then generated. Figure 8 shows one page from this listing. Since no part type changes have been to this point the predicted chassis AD failure rate was of course 25.13 per cent per thousand hours, the same as the total of the failure rates of the four circuit code groups.

The chassis was then designed. The design of a chassis often dictates changes in some part types because of space, weight, or availability considerations. Generally after the chassis has been designed all part types are known. At this point the chassis AD parts list was modified. In some instances part type changes were made. A more formal type of part designation was substituted for the development phase description. Where type and rating changes were made, the failure rate was changed. The IBM deck was modified accordingly. This is "prediction and control deck number 2" in Figure 1. A new parts and reliability listing was generated from this deck. Figure 9 shows one page from this listing. The listing can be furnished on multilith masters or vellums by the data handling facility. The vellums can be incorporated on other drawings or released separately to manufacturing. It is also possible to release a duplicate deck of IBM cards.

The prototype production failure rate estimate was compared to the prediction at the completion of the circuit design. The new chassis AD failure was 25.35 per cent per thousand hours.

If the failure rate had been significantly higher than the circuit design phase failure rate and the reliability specifications had not been met, action would have to be taken at this early stage.

* Since numeric designations had been used to identify the circuit code groups, it was necessary to use letter designations (AD) to identify the chassis.

The action would have consisted of either partial redesign of the circuitry, the chassis layout, or a search for better components or possibly all three.

"Prediction and control deck number 2" would be modified as dictated by the performance of the prototype production chassis during the test and evaluation period. This modified deck would be "prediction and control deck number 3", the factory production parts list and reliability deck. See Figure 1.

Measured Reliability

As shown in Figure 1 there should be an "observed failure rate" card deck associated with each "prediction and control" deck. As in the "prediction and control" deck, there would be a card for each component. The observed or measured failure rate information for a component would be punched on one card. Periodically the "prediction and control" deck would be compared in regard to failure rate with the associated "observed failure rate" deck. The comparison can be made on the basis of what is considered a significant difference. This will depend on the confidence in the estimated failure rates and the statistics connected with the observed failure rates.

The "observed failure rate" deck can not be generated directly from failure reports. There will have to be an intermediate "failure report deck". The "observed failure rate" deck will be generated by suming the component operating time and failure information in the "failure report deck".

PART TWO

This section describes extensions of the system.

When a circuit design is completed and an IBM deck has been punched, other IBM listings, in addition to the conventional parts listing, might be requested such as: over-all listing by component type with a summation of failure rates for each type or listing by derating to indicate which components are not adequately derated. These same listings would be useful when the final prototype chassis parts list is on IBM cards. In addition, a listing by similar parts can be obtained for the purchasing section.

Sufficient information is stored and available for the reliability section to assess reliability, locate danger areas, and make recommendations for improving reliability. The following listings are typical of what might be useful for this application: 1. Development, prototype production, and field phase listing by failure; indicating: component failing, stress, derating, running time, turn on and turn off cycles, and type of failure. 2. Accumulated failures vs. turn on and turn off cycles (all phases). 4. Accumulated failures vs. type of failure (all phases). 5. Failure vs. particular field site and ad infinitum as necessary to correlate failures with the causes.

The memory facility of a computer could be utilized to store the following information: approved parts, approved vendors, and part failure rates vs. stress and environment. The circuit designer would then have to enter on the card punching sheet only the following information: Circuit symbol number, nominal value, permissible tolerance, stress, and operating temperature. This would eliminate a tedious and time consuming phase for the circuit design engineer and also reduce errors.

PART THREE

The data handling described in this paper was done at an EAM (Electric Accounting Machine) facility. The procedure was as follows: the circuit design engineer listed the component information on a punching sheet. The punching sheet went through a control group in the circuit design section to the key punch operator. The operator punched an IBM card for each component. A second operator verified the punched card. When all the cards for a particular circuit group were punched and verified the deck was placed in a sorter which arranged the cards in the desired sequence. Initially, the cards were ordered by component type and then the groups of components were arranged numerically. This was the conventional parts list format. The deck of cards was then placed in the printer-tabulator which printed a listing of the components and totaled the failure rates. A listing can be made on vellum, multilith master, or paper. If this initial listing is the only one required and if the circuit designer lists the component in the desired order, it would be cheaper and more efficient to employ a typist to generate the part description and reliability lists rather than to use the EAM facility. If different type listings and/or if frequent up dating are required the IBM card system should be used. The IBM card system is quite flexible and, therefore, lends itself to many uses, some of which are unanticipated at the inception of a program.

The next phase was the generation of the prototype production parts lists. This was the phase where the parts lists for several circuit groups were combined to make up a chassis parts list. This was done with the aid of a transposition list of development schematic circuit symbols to chassis circuit symbols. The data processing facility punched a transposition deck, one card for each circuit symbol, and then used the development circuit decks to automatically punch the cards

for the chassis parts list deck. The cards were then machine ordered and a chassis part listing was printed. This listing included all the stress, derating, and failure rate information included in the development listings. The generation of a chassis parts list from the development circuit parts list by means of a transposition list is a definite application for the EAM facility.

The more elaborate schemes described in part two such as the automatic "table look up" of approved parts, approved vendors, and failure rate vs. temperature would require an EDPM (Electronic Data Processing Machine) system. In order to justify such a program, there must be a general application in many areas of a large company. Alternately such a system might service many companies.

CONCLUSION

An IBM card data handling system has been described by means of which one can predict, control and measure the reliability of an electronics unit. It would be impractical if not impossible to implement such a program without using machine type data handling.

Potential users of such a system should examine critically the relative complexity of their program to determine whether EAM, EPDM or manual data handling is required.

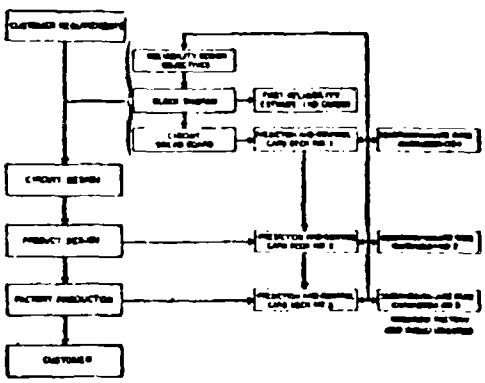


Figure 1

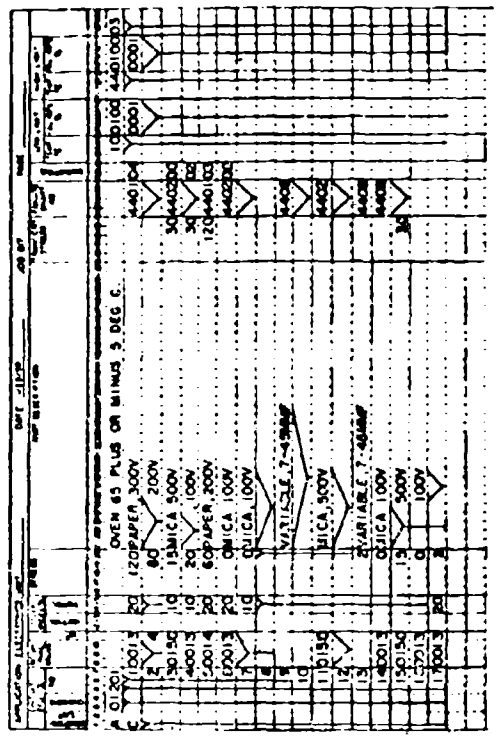


Figure 2

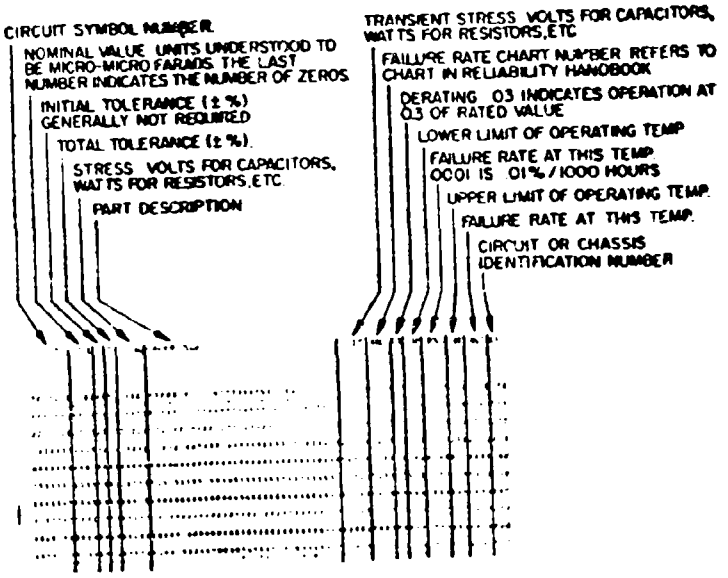


Figure 3

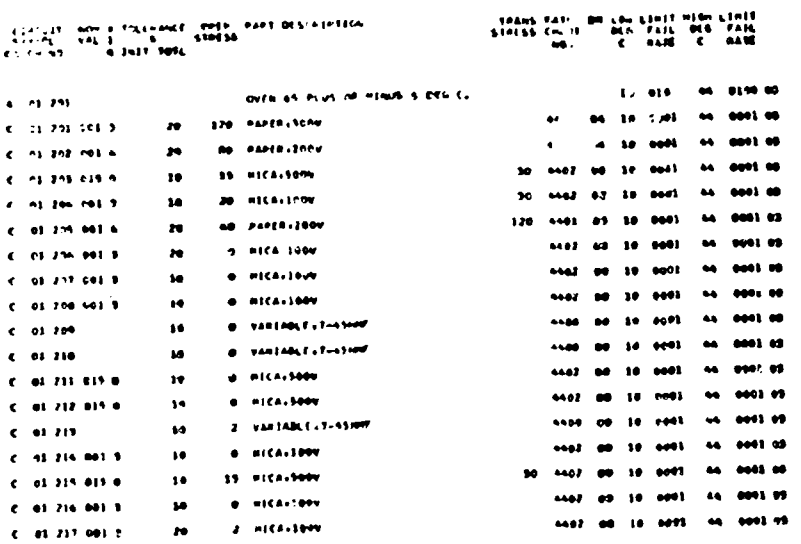


Figure 4

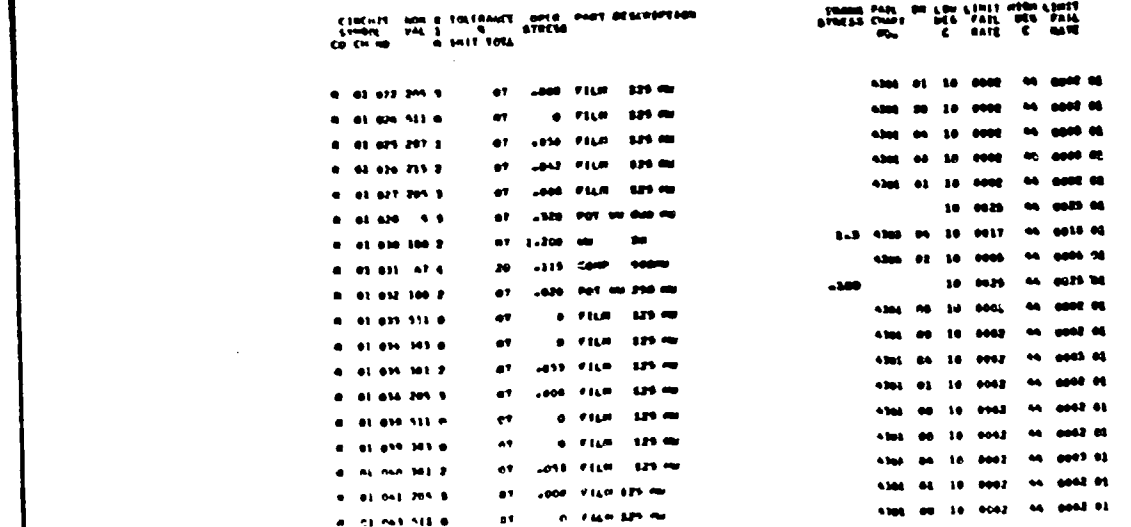


Figure 5

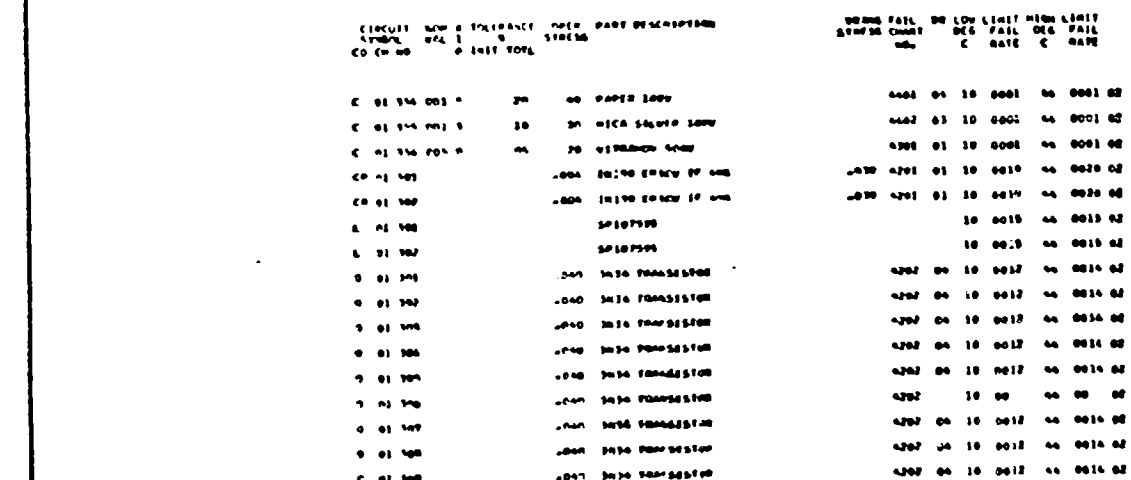


Figure 6

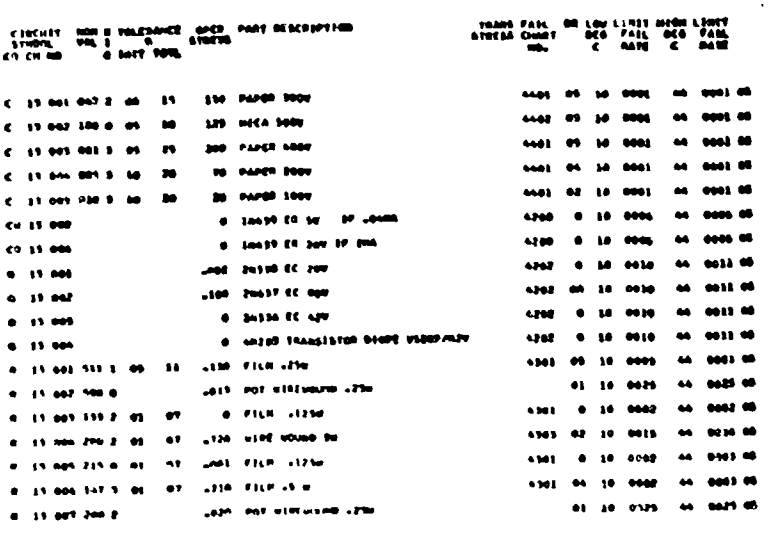


Figure 7

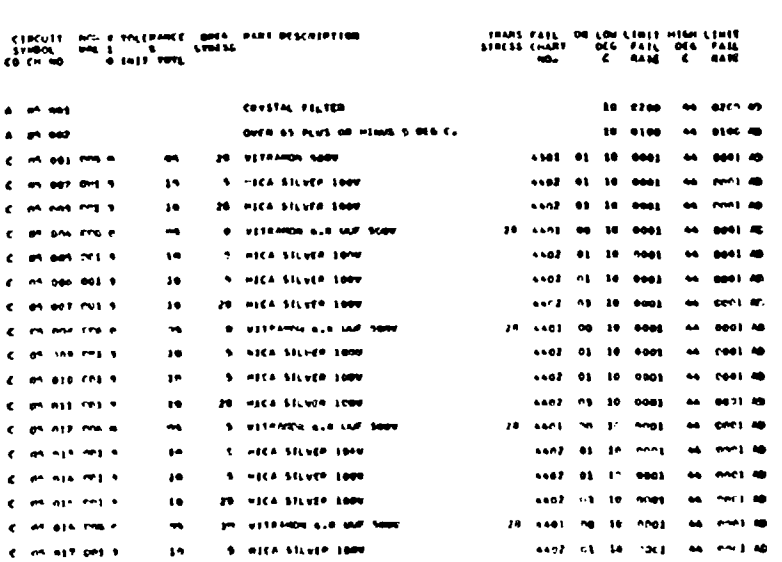


Figure 8

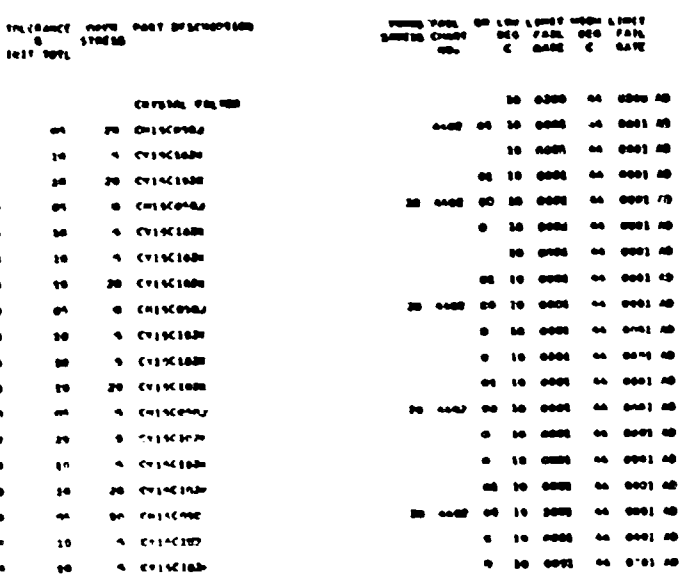


Figure 9

SYSTEM IMPROVEMENT THROUGH FAILURE EFFECT AND A RELIABILITY MODEL

George E. Unruh
Paramount Electronics, Inc.
Hicksville, L.I., N.Y.

Summary. The principles of nature derived from the perennial philosophy are used to develop a reliability program in which the main features are a failure effect analysis and a reliability model.

Introduction

In a previous paper we discussed the distinction of the four causes of artifical things.[3] In this paper we shall discuss the development of some aspects of a reliability program, beginning with the principles previously discussed.

Functional Analysis of System.

The final cause of things is said to be the cause of causes.[1] In artifical things this is easy to see, since the purpose of a machine will cause its design to be such and such - and the material to be of a certain kind. Even the methods of manufacture will be fashioned according to the purpose of the machine. Thus the final cause gives the character to all aspects of the artifical things.

Accordingly we must begin our analysis of a system by a study of the purpose of the system. An example of this type of analysis is sketched here in figure 1, as taken from a surveillance system. This type of analysis is always undertaken in designing a system. It should also be done in preparing a reliability program.

It is possible, after preparing the division of the system into functional parts, to assign complexity figures and to apportion the complexities among the divisions. However the procedure is mathematical and should be considered after a physical examination.

Failure Effect Analysis.

Since we have begun by an examination of the end purpose, we should continue to formulate our program along these lines and prepare a more detailed consideration of the system parts which will, by failing, contribute to system failure - that is failure of the system, to accomplish its end purpose.

Some have recommended that this consideration be merely the preparation of a list of critical parts, leading to a review of the parts themselves. However this procedure omits a consideration of the system design as ordered to successful accomplishment of purpose. A procedure which considers the purpose of the system and which arranges the parts in causal relationship, showing the degrees of criticality is a failure effect analysis.

Analysis Procedure.

A failure effect analysis of every circuit or functional part of the system naming type of part failure, the proximate effect and the remote effect of the failure in the operation of the system will disclose ways to improve the system by part improvement, redundancy or redesign.

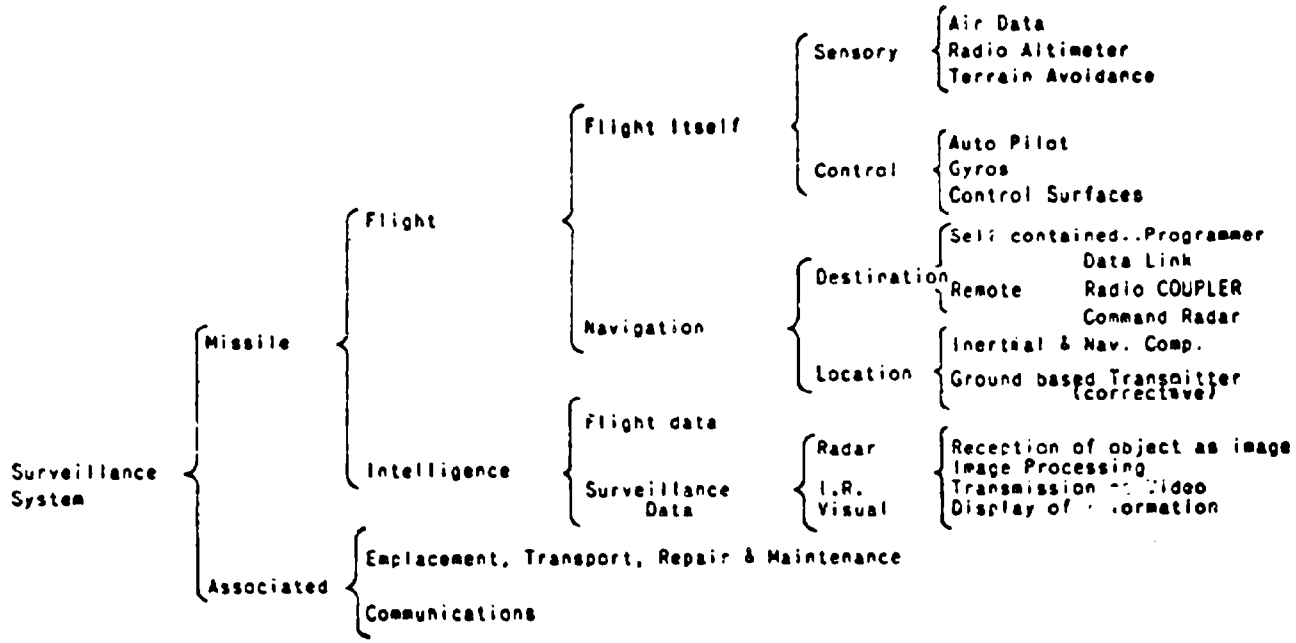
A sample of this failure effect analysis, taken from a radar system is shown in figure 2.

The type of failures chosen for consideration should be:
(1) The distortion of the output signal of the circuit and;
(2) The change in value of parameters in the circuit.
The degree of detail of this analysis and the extent to which this is carried depend on the purpose of the system. The type of failure effect analysis described here was chosen because the end function of the system considered is rather simple, being radar transmission and display of the radar picture.

Benefits of Failure Effect Analysis.

The benefits of a failure effect analysis are as follows:

(1) Selection of critical and non-criti-



FUNCTIONAL DIVISION OF A SURVEILLANCE SYSTEM

Figure 1.

NAME OF CIRCUIT	TYPE OF FAILURE	PROXIMATE EFFECT	REMOTE EFFECT
Master Multivibrator	Asymmetry	Time displacement of negative Master trigger	Critical distortion of video displays.
Master Multivibrator	No signal	No positive or negative trigger	No radar transmission
Range Marker	Absence of trigger pulse.	No range markers.	Non-critical absence of marker in display.

Failure Effect Analysis Form

Figure 2.

cal circuits.
(2). Increased knowledge of system design making possible recommendations for redundancy, simplification or circuit improvement.
(3) Preparation of an improved maintenance program and technician training course.

The failure e fect analysis will immediately disclose which circuits are necessary to system operation. Of course decisions will have to be made evaluating the importance of various system functions and subordinating some functions to others in the light of the accomplishment of the final purpose of the system. This is a valuable by-product of this method of analysis, and useful to system engineering groups.

The circuits whose failure will cause system failure as regards attainment of the operational purposes of the system must by called critical.

This is the proper use of the word "critical" in reliability studies, since the "criteria" for judging a system or its parts is satisfactory operation and attainment of goals in the field.

The Reliability Model.

Thus far we have described studies of the system as regards physical effects signals, circuit failures, etc. It is necessary also to consider the mathematical relations of probability and to make a schematic showing the probabilities of failure which act in series and those which are in parallel, or redundant. Such a schematic may be called a Reliability Model, or better a Probability Model. A Probability Model of the entire system as well as important parts of the system should be made.

In a Probability Model the relation between failure rates in a system are shown schematically. Failure rates of components whose functions are causally related are usually added. A Multivibrator produces a waveform which is used by a differentiator to produce two triggers, one of which is used by a transmitter to turn the transmitted pulse on. This causality becomes a sort of alternating final and efficient causality.

Series and Parallel Probabilities.

These failure rates should be schematically represented as being in series, as resistors are in series in a network. As the total resistance is the sum of the resistances in series, so the total failure rate is the sum of the individual failure rates. Thus the analogy, which is one of proportionality.

For two components causally related in the way mentioned, with reliabilities R_1 and R_2, the resultant reliability R is;

$$R = R_1 \cdot R_2$$

If f_1 and f_2 are the failure rates

$$R = R_1 \cdot R_2 = e^{-f_1 t} e^{-f_2 t} = e^{-(f_1+f_2)t}$$

Thus the resultant failure rate is:

$$f = f_1 + f_2$$

The failure rates of components which are redundant or whose functions are schematically represented as being in parallel. The failure rates of redundant components is determined by

considering the unreliability or probability of failure of the combinations.[2]

$$Q = Q_1 \cdot Q_2 = (1 - R_1)\ (1 - R_2)$$

or

$$R = 1-Q=1-Q_1 \cdot Q_2=1-(1-R_1)(1-R_2)$$
$$R = R_1+R_2-R_1R_2=e^{-f_1t}+e^{-f_1t}-e^{-(f_1+f_2)t}$$

and the mean time between failures is:

$$M = \frac{1}{f_1} + \frac{1}{f_2} - \frac{1}{f_1 + f_2}$$

If the failure rates are equal

$$M = \frac{3}{2} \cdot \frac{1}{f_1}$$

or

$$f = \frac{1}{M} = \frac{2}{3} f_1$$

For three components in the equal failure rates in parallel,

$$R = 1-Q_1 \cdot Q_2 \cdot Q_3 = 1-Q_1^3 = 1-(1-R_1)^3$$

$$M = \frac{1}{f_1} + \frac{1}{2f_1} + \frac{1}{3f_1}$$

For four components,

$$R = 1 - (1 - R_1)^4$$

$$M = \frac{1}{f_1} + \frac{1}{2f_1} + \frac{1}{3f_1} + \frac{1}{4f_1}$$

for n components.

$$R = 1-(1-R_1)^n$$

$$M = \frac{1}{f_1} + \frac{1}{2f_1} + \frac{1}{3f_1} + \text{----} + \frac{1}{nf_1}$$

The formation of a reliability model, or a probability model is accomplished by insight into the system itself and knowledge of the relation and interdependence of functions.

Any lack of such understanding is immediately reflected in an inadequate or untrue model.

Example From Multiplex System.

An example of the usefulness of this model is taken from the case history of a multiplex system. In this system 'x' different frequencies, are selected at random, to make possible 'xy' possible sum frequencies occurring at random. It is essential to the operation of the system that no one frequency be transmitted longer than a certain very short time.

Each of the oscillator stages of this multiplexer are turned off by a double switch, that is by a diode gate and a gated amplifier, both of which were arranged so that the probability of failing in the on position was greater than failing in the off position. The reliability model immediately showed that the probability of an oscillator remaining on was dangerously high. Accordingly it was recommended to the engineering design section that the bias voltage be set to fail safe in the "off" condition, so that the probability of failing off would be higher than the probability of failing on. The reliability of the system, an important link in the defense network was increased notably by this comparatively simple improvement. The reliability model now showed, after the improvement, that the probabilities of failure of the switches were in parallel for the on position, but in series for the off position.

Example from Sealed Unit.

Another common problem that may be solved using the reliability model is finding the number of units required for a sealed system in which no replacements can be made. A package of power diodes is required for a power supply. The probabilities of failure are in parallel. A given number are required for successful operation. How many more are required to give the probability R_s that the package will not have to be replaced for the life of the equipment. To solve this problem we use the binomial expansion.

$$(R+Q)^n = R^n + nR^{n-1}Q + \frac{n(n-1)}{2!}R^{n-2}Q^2 +$$
$$+ \text{- - - -} Q^n = 1$$

R^n is the probability that none of the units will fail.

$nR^{n-1}Q$ is the probability one of the units will fail, etc.

Knowing the failure rate of each of the several units, the minimum number of units required for successful operation even for the briefest time, and the required overall reliability of the system, it is possible to calculate the number of additional units, beyond the minimum, to be built into the sealed system.

By using a failure effect analysis to discover critical parts and circuits and the formation of a reliability model, the project engineer and his aides can judiciously select the circuits to which

redundancy may be most profitably applied. The use of diagrams and schematics will make the project easier and again prove the adage: "A picture is worth a thousand words".

Conclusion

Thus we have given an example of reasoning from an abstract understanding of the principles of the nature of an artificial product to the development of a program for improving the reliability of the product.

References:

1. St. Thomas Aquinas: Commentary on the Physics of Aristotle. Book 2.

2. Boeing Airplane Company: Reliability Handbook D6-2770, Part I. Defining, Calculating and Designing Reliability.

3. G.E. Unruh: An Integrated Program for Reliability. Sixth National Symposium on Reliability and Quality Control in Electronics, Jan. 11-13, 1960.

FERROMAGNETIC ANTENNAS

Ronald Row
American Electronic Laboratories, Inc.
Colmar, Pennsylvania

INTRODUCTION

In considering the efficient transmission and reception of electromagnetic energy in the HF and VHF ranges of the frequency spectrum, it has been generally accepted that resonant or near resonant antennas be used. It is obvious that the corresponding size and weight of such designs are not compatible with the normal military requirements of mobility and portability.

The usual alternative is the use of electrically short structures designed to "fit" mechanically, but with much to be desired from a radiating or receiving efficiency point of view. Efficiency, in general, is related to the ratio of radiation resistance to loss resistance (r_r/r_l). Techniques of reducing loss resistance have been reported previously[1, 2] and are in general characterized by complex ground systems, high Q reactive tuning components, and use of highly conductive metallic radiating elements. A more recent interest[3, 4] has been directed toward methods of increasing the radiation resistance component by dielectric and/or magnetic loading of the radiating element(s) ------ the subject of this paper.

HISTORY

Antennas embedded in a dielectric or ferromagnetic medium have been investigated by Herman[3] and Folk[4]. Their principal aim in utilizing a medium other than air was the effective increase in electrical dimension of the radiator and the accompanying transformation of the series radiation resistance. Herman has reported that the magnitude of the resistance transformation is directly proportional to the intrinsic impedance of the surrounding medium. His experimental work confirms this theory to the extent that apparent increases in antenna resistance of 30 to 40:1 were noted when a biconical antenna was immersed in a spherical ferrite medium of diameter equal to the total antenna length (Figure 1). It is important to note, however, that the greater part of the increase in antenna resistance was attributed to the loss resistance of the ferrite material with only a small, relatively insignificant increase in radiation resistance.

Folk has considered the case where the radiating element is physically "short", but electrically "long" with respect to the wavelength of the surrounding medium. His calculations indicate that the first resonance* radiation resistance is essentially equal to that of an air antenna of the same length, with some small variation to be expected as a function of the relative permeability and permittivity of the medium.

*First resonance is defined as the first frequency at which the reactance goes to zero when moving from $\omega = 0$ —►

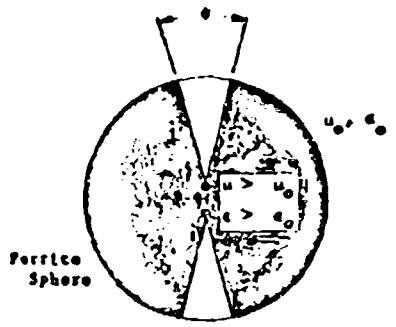


Figure 1

Biconical Antenna Immersed in a Ferrite Sphere

CURRENT WORK

Lipkin has chosen to consider the problem in light of the contribution of the magnetizable material to the radiating current moment of the small electric antenna.

The far-field amplitudes produced by an electrically small antenna are proportional to its current moment, defined for an electric antenna as the volume integral of its effective conduction-current density.

The effective conduction current density consists of three parts, given by the expression:

$$J + \text{curl } M + \frac{1}{c}\frac{\partial P}{\partial t}$$

where J = the vector density of current in conducting bodies

M = the vector magnetic moment per unit volume in magnetic bodies

P = the vector polarization in dielectric bodies

Curl M is a fictitious conduction-current density attributable to the presence of magnetic bodies. That is, if all magnetic parts were removed from the antenna, the conduction-current density distribution, curl M, if supplied by external means, would maintain unchanged the electromagnetic field existing everywhere outside of the former locations of the magnetic parts. Similarly the term, $1/c \cdot \partial P/\partial t$, or density of polarization current, is a fictitious conduction-current density attributable to the presence of all the dielectric parts of the antenna.

The effect of using magnetizable material in

an electric antenna may be analyzed by computing the volume integral of curl M; thus, the magnetic contribution to current moment is

$$K = \int \text{curl } M \, dx \, dy \, dz.$$

This integration can be separated into two steps. The first is an area integration over a slice taken through the antenna, and the second is an integration performed as the slice is displaced parallel to itself through the entire length of the antenna. Thus:

$$\text{Let } F = \int \text{curl } M \, dx \, dy.$$

$$\text{Then } K = \int F \, dz.$$

Here, x, y, and z are orthogonal cartesian coordinates oriented such that the slices taken through the antenna lie in planes of constant z. The areas over which F is defined may be taken to extend beyond the boundary of the antenna by any definite amount but not so far as to enclose the receiving antenna. Consider the z component of F:

$$F_z = \int (\text{curl } M)_z \, dx \, dy.$$

By Stoke's theorem, this area integral can be reduced to a line integral around a contour C which bounds the area of integration:

$$F_z = \int_C M \cdot dl$$

Mathematical requirements of continuity do not basically restrict the validity of the reduction of F_z thus accomplished. Due to the choice of the area for integration, the bounding contour C will lie entirely outside of the antenna, in free space. Along the contour C there will consequently be no electric or magnetic polarizations to consider, and in particular M = 0 on C. Thus $F_z = 0$, identically. This in turn requires that $K_z = 0$. Similar arguments show that K_x and K_y also vanish. This completely general result can be stated as follows:

In principle, the utilization of dielectric or magnetic materials enclosing an electric dipole cannot contribute to the radiating current moment of the antenna.

In anticipation of the question as to what the presence of magnetizable material can accomplish at all, it can be stated that the presence of such material can profoundly affect the terminal impedance of a small electric antenna. This change, however, is for the most part reactive; any significant change in real part can be attributed to losses in the magnetic material, with exception to the small increase in radiation resistance resulting from a change in current distribution, as explained in a later section.

A prime example is an antenna in which inductive loading provided by the magnetic material tunes out all or a substantial part of the capacitive reactance which the antenna normally presents to the load or to the driving source. A ferromagnetic antenna which has zero reactance at a given frequency by virtue of the inductive effect of the magnetic material it contains may be said to be self-resonant at that frequency. If the frequencies of self-resonance are controlled by saturation of the magnetic material, the ferromagnetic antenna may be said to be "self-tunable".

In addition to producing self-resonance in a small electric antenna at frequencies much lower than would be possible in air, the presence of magnetizable material exercises a small degree of control over the radiation resistance by virtue of its effect on the antenna current distribution on the radiating conductors.

For example, a small antenna which would have a linearly tapered current distribution in air will, when made self-resonant by the use of magnetic material, have a sinusoidal or near sinusoidal current distribution whose associated radiation resistance may approach four times that associated with the linear current distribution.

This is more readily understood by considering the effect of current distribution on radiation resistance.

Case I - Linear Current Distribution

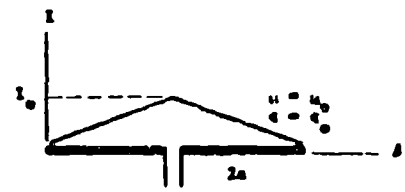


Figure 2

Current Distribution on a Dipole Antenna of Length $2a \ll \lambda_{air}$ (Medium - Air)

In this case, the current distribution will be almost linear, tapering to zero at the antenna ends. If the antenna terminal current is I_o, the current moment (area under the curve) of the antenna which determines the useful radiated power will have the value aI_o. The radiation resistance will assume a value proportional to the square of the current moment, or:

$$r_r \propto (aI_o)^2$$

Case II - Sinusoidal Current Distribution

The sinusoidal distribution, given to exist on structures of resonant length produces a current moment given by:

$$\text{Current Moment} = \int_0^{2a} I_o \sin \delta \, dl = 2I_o$$

$$\text{where} \quad 2a = \frac{\lambda_M}{2} \text{ or } \pi \text{ radians}$$

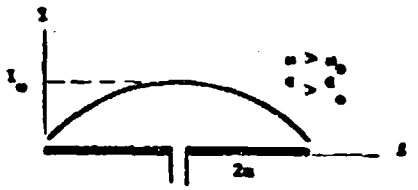


Figure 3

Current Distribution on a Dipole Antenna of Length 2a; $\frac{\lambda_M \text{(Medium Wavelength)}}{2} = 2a \ll \lambda_{air}$

(Medium - Ferrite)

The increase in current moment in Case II over Case I is then,

$$\frac{2I_o}{aI_o} = \frac{2}{\delta} = \frac{4}{\pi}$$

This corresponds to an increase in radiation resistance, or useful radiated power of $16/\pi^2$.

Case III - Constant Current Distribution

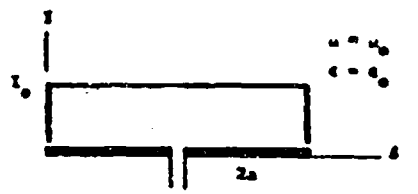


Figure 4

Current Distribution as May Appear on a Heavily End Loaded Dipole. $2a \ll \lambda_{air}$ (Medium - Air)

Capacitive top or end loading, if complete will produce a constant distribution of current on the conductors up to the loading elements, where the current sharply tapers to zero. Within a reasonable degree of accuracy, this distribution determines a current moment equal to $2aI_o$ and useful radiated power proportional to $(2aI_o)^2$

Normalizing the current moment and radiation resistance of Case I, one can see (Table I) the variation to be expected as a function of the geometry of the current distribution.

EXPERIMENTAL DEVELOPMENTS

Experimental verification of the theory developed by Lipkin was accomplished by comparing the transmitted power of a base tuned eight foot air monopole (Figure 5) to that of a ferrite-sheathed monopole of the same physical height (Figure 6). Measurements showed the ferrite monopole to have a first resonance at 5.235 mc. The base tuning inductance was so chosen to resonate the air monopole at this frequency. After proper compensation and correction for transmission line mismatch, circuit Q, etc., the far field electric intensity generated via the ferrite monopole was measured to be 1.85 times that of the air monopole. The corresponding effective increase in radiation resistance was $(1.85)^2 = 3.4$.

The fact that the measured ratio of radiation resistances exceeded 1.6 can be taken to indicate that the current on the ferrite monopole was distributed in a flattened sinusoidal manner. This distribution could presumably be explained by a detailed boundary-value theory of the ferrite-sheathed monopole.

Apparent success in the above precipitated the development of a full scale model, tunable throughout the range of 8 kc to 3.5 mc as dictated by contract requirements.

The choice of antenna type was based upon the requirement of maximum radiation efficiency for a given physical size.

Four basic antenna types were considered.

A. The Electric Dipole Antenna (Figure 7a). It is cylindrical and has a total length (L_1) much greater than its diameter (D_1).

B. The Solenoid Antenna (Figure 7b). A loop antenna wound in the form of a solenoid whose axial length (L_2) is much greater than its diameter (D_2).

C. The Flat Loop Antenna (Figure 7c). It is circular and has a mean major diameter (D_3) much larger than its minor diameter (L_3).

D. The Capacitor Antenna (Figure 7d). A heavily end-loaded electric dipole antenna, whose end-plate diameter (D_4) is much larger than the separation (L_4) of these plates.

If reasonably small, each of these antennas will appear as almost a pure reactance to its driving source at low radio frequencies. The radiation resistance of these antennas will, in general, form only an extremely small fraction of the real part of the antenna input impedance. This real part is due almost exclusively to losses. The ratio of the power lost by the antenna to the reactive power fed to it (or what is nearly the same thing, in the low-efficiency case, the ratio of the real part of the input impedance to the reactive part) defines a ratio 1/Q for a reactive antenna. It will be assumed here that the antennas are tuned. The factor Q can be taken to include the losses of the tuning reactor as well as those of the antenna.

The ratio of useful radiated power to reactive power equals the ratio of series-equivalent radiation resistance to antenna reactance, or r_r/x. The ratio of useful radiated power to lost power will then be r_r/x divided by 1/Q or Qr_r/x.

This can be taken to define a power efficiency "η", for the antenna as a radiator. The "efficiency" so defined is in close accord with the usual expression for efficiency whenever it is much less than unity. The quantities r_r and x can be calculated readily for the four antenna types under consideration.

The series-equivalent radiation resistances of the four antenna types are:

Thin Dipole Antenna

$$r_{r1} = \left[\frac{2}{3}\pi^2 c\ 10^{-9} \frac{L_1^2}{\lambda^2}\right] \text{ ohms}$$

Solenoid Antenna

$$r_{r2} = \left[\frac{2}{3}\pi^6 c\ 10^{-9} \frac{D_2^4 N_2^2}{\lambda^4}\right] \text{ ohms}$$

Flat Loop Antenna

$$r_{r3} = \left[\frac{2}{3}\pi^6 c\ 10^{-9} \frac{D_3^4 N_3^2}{\lambda^4}\right] \text{ ohms}$$

Capacitor Antenna (for radiation polarized at right angles to the plates)

$$r_{r4} = \left[\frac{8}{3}\pi^2 c\ 10^{-9} \frac{L_4^2}{\lambda^2}\right] \text{ ohms}$$

Using readily available approximate formulas for the capacitance or inductance of the respective antennas, their reactances can be put in the form:

Thin Dipole Antenna

$$x_1 = \left[\frac{2}{\pi} c\ 10^{-9} \frac{\lambda}{L_1} \operatorname{Ln}\left(\frac{2}{e}\frac{L_1}{D_1}\right)\right] \text{ ohms}$$

Solenoid Antenna

$$x_2 = \left[2\pi^3 c\ 10^{-9} \frac{N_2^2 D_2^2}{\lambda L_2}\right] \text{ ohms}$$

Flat Loop Antenna

$$x_3 = \left[4\pi^2 c\ 10^{-9} \frac{N_3^2 D_3}{\lambda} \operatorname{Ln}\left(\frac{8}{e^2}\frac{D_3}{L_3}\right)\right] \text{ ohms}$$

Capacitor Antenna

$$x_4 = \left[\frac{8}{\pi} c\ 10^{-9} \frac{\lambda L_4}{D_4^2}\right] \text{ ohms}$$

In the equations expressing r_{rn} and x_n, λ is the free-space wavelength at the radio frequency under consideration; c is the velocity of light expressed in cm/sec; $e = 2.718\ldots$; and N_2 and N_3 are the number of turns used in winding the solenoid and flat loop antennas, respectively.

Using the above results for radiation resistance and reactance, the following antenna radiation efficiencies are calculated, using $\eta_n = Q_n r_{rn}/x_n$:

Thin Dipole Antenna

$$\eta_1 = \left[\frac{\pi^3}{3} Q_1 \frac{L_1^3}{\lambda^3} \frac{1}{\operatorname{Ln}\left(\frac{2}{e}\frac{L_1}{D_1}\right)}\right]$$

Solenoid Antenna

$$\eta_2 = \left[\frac{\pi^3}{3} Q_2 \frac{D_2^2 L_2}{\lambda^3}\right]$$

Flat Loop Antenna

$$\eta_3 = \left[\frac{\pi^4}{6} Q_3 \frac{D_3^3}{\lambda^3} \frac{1}{\operatorname{Ln}\left(\frac{8}{e^2}\frac{D_3}{L_3}\right)}\right]$$

Capacitor Antenna

$$\eta_4 = \left[\frac{\pi^3}{3} Q_4 \frac{D_4^2 L_4}{\lambda^3}\right]$$

Examination of the four efficiency formulas indicates that where Q's can be assumed equal, geometrical shape considerations come to the fore. All four antenna types under consideration have circular cylindrical symmetry, with overall physical dimensions "L" in the axial direction and "D" at right angles to that direction. The four efficiency formulas involve the following quantities having dimensions of volume: L^3, LD^2, and D^3.

The obvious choice, the electric dipole, involving the term L^3, provides a maximum radiation efficiency in given (minimum) physical dimensions. When very large structures are required (low frequency operation) the dipole is normally replaced by a monopole (stub) over a ground plane.

A 30-foot vertical monopole was constructed as shown in Figure 8. The ferrite enclosure was abandoned in favor of a technique employing closed ferrite-core inductors (Figure 9a) distributed at equal increments along the radiating element. This technique offered provisions for biasing the ferrite core inductors by application of a dc magnetizing current to obtain a tuning capability (permeability tuning).

Shortly after the inception of this program it was realized that the core losses in the ferrite material were so great, due to the closed magnetic path, that the advantage to be gained by distributed tuning would be insignificant compared to additional loss resistance present in the tuning circuit.

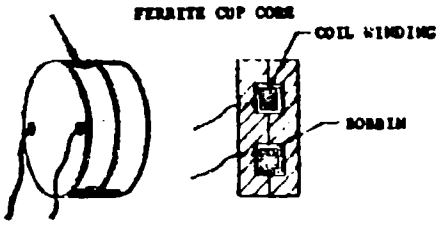


Figure 9a. CLOSED FERRITE CORE INDUCTOR

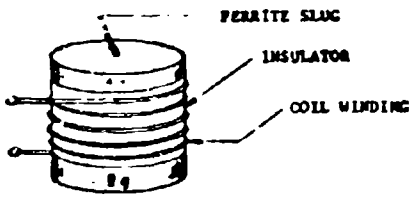


Figure 9b. OPEN FERRITE CORE INDUCTOR

At this point, further consideration was given to base tuning the receiving antenna; here low loss (high Q) variable capacitors could be used in lieu of the lossy ferrite core inductors.

Open ferrite core inductors (Figure 9b) carefully designed for maximum Q factor were combined with variable air capacitors (Figure 10) to tune out the antenna capacity. Q factors of 1000 or more were realizable with the open core inductors. Circuit bandwidth was controlled by variation of the parameters L_t, C_t, and R_{L_t}.

This tuning technique offered the advantage of voltage step-up to the grid of the first stage of the receiver, eliminating the need for lossy, difficult to build, step-up transformers. A cathode follower (impedance transformer) was introduced between the tuning circuit and the coaxial feed cable to eliminate the resonant effect of the long transmission line.

Bench tests were performed comparing the sensitivity (signal to noise ratio) of the receiving system ---- with, and without the tuning coupler. Improvements on the order of 25 db in S/N ratio were noted when the coupler was introduced into the receiving system.

It is noted, however, that the effect of ambient noise (prevalent at low frequencies) on the S/N ratio of the receiving system was not investigated due to the complexity of the measurement problem. To ascertain its effect the equipment will undergo a complete field evaluation under the cognizance of the United States Army Signal Research and Development Laboratories.

CONCLUSIONS

The presence of ferromagnetic material surrounding a small antenna of any type does not contribute materially to an increase in radiation efficiency. The presence of the material can indirectly affect the radiation resistance by causing the distribution of current on its conducting elements to differ from that which would have existed in air. This effect on radiation resistance is potentially limited to a 4:1 improvement and will not exceed effects obtainable by conventional methods of altering the antenna current distribution, e.g., by top loading.

The principal advantage to be gained by immersion of a small antenna in a ferromagnetic medium is its ability to provide reactance cancellation; even here the merits must be considered in light of the huge losses associated with its use.

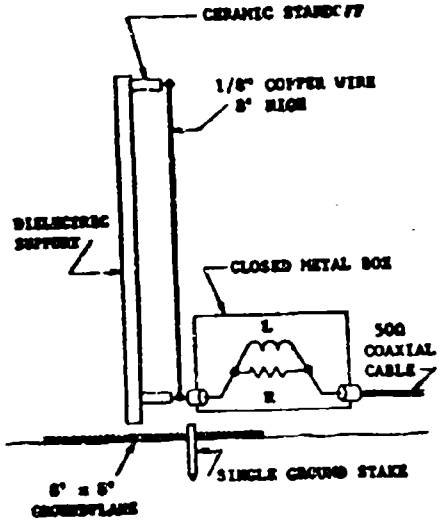
One must, therefore, be satisfied by optimizing present techniques (as was accomplished on this contract) until such time as new methods or arts are devised.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Mr. Daniel Lipkin, American Electronic Laboratories, (now with Republic Aircraft Company) for his contribution to the theoretical aspect of this work, and to the United States Army Signal Research and Development Laboratories for their guidance and support.

REFERENCES

1. F.E. Terman, "Radio Engineers' Handbook", McGraw-Hill Book Company, 1943; pp 74-109.
2. G. Brown, R. Lewis and J. Epstein, "Ground Systems as a Factor in Antenna Efficiency", Proceedings of the IRE, June 1937.
3. J. Herman, "Thin Wire Loop and Thin Biconical Antennas in Finite Spherical Media", Diamond Ordnance Fuze Lab., Washington, D.C., Tech. Report 462; May 1, 1957.
4. C. Polk, "Resonance and Supergain Effects in Loaded Antennas", IRE Transactions on Antennas and Propagation, Vol. AP-7, pp 414-423; December 1959.



L = Tuning Inductor

R = Resistor to lower "Q" of tuning inductor, increasing its equivalent series resistance

During gain measurement at 5.235 mc, L = 34.1μh, R = 270 KΩ

Figure 5. 8-FOOT AIR ANTENNA

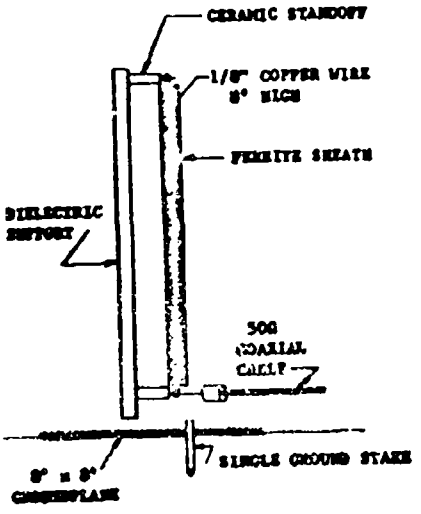


Figure 6. 8-FOOT FERRITE-SHEATHED ANTENNA

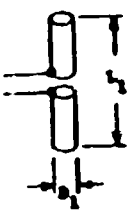


Figure 7(a)
Dipole Antenna

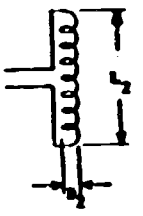


Figure 7(b)
Solenoid Antenna

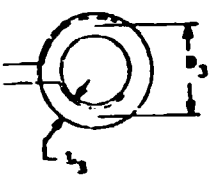


Figure 7(c)
"Flat" Loop Antenna

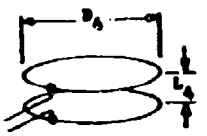
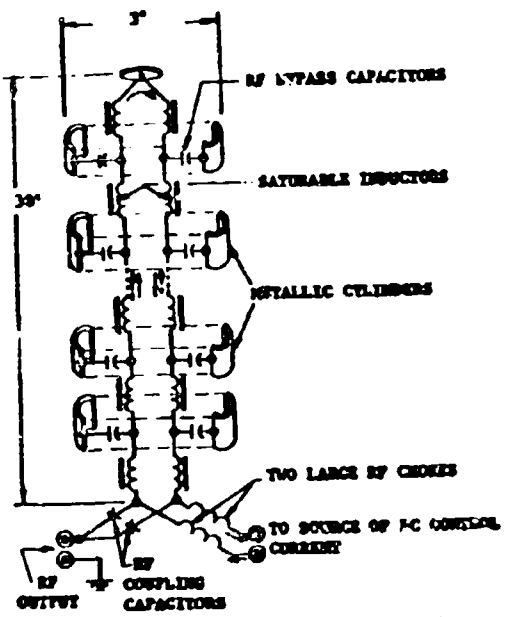


Figure 7(d)
Capacitor Antenna

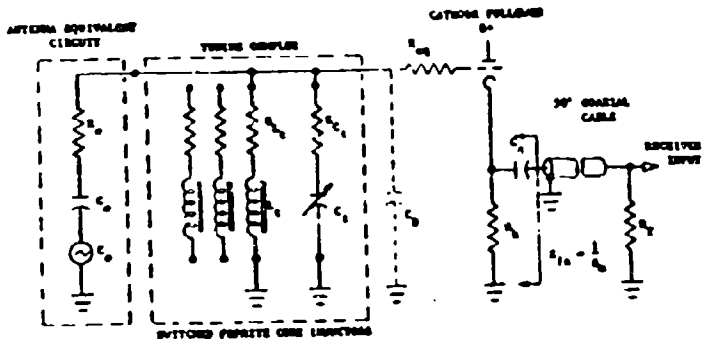


Solid arrows indicate control currents.
Dashed arrows indicate RF currents.

Figure 8. FERROMAGNETIC ANTENNA (TOROID)

ANTENNA	CURRENT DISTRIBUTION	CURRENT FACTOR	RADIATION RESISTANCE
Dipole, $2a \ll \lambda_{air}$ (Medium - Air)	Linear	1.00	1.00
Dipole, $2a = \frac{\lambda}{2}$ (Medium - Ferrite)	Sinusoidal	1.27	1.62
Dipole, $2a \ll \lambda_{air}$ Heavily cap loaded (Medium - Air)	Constant	2.00	4.00

Table 1. CURRENT MOMENT AND RADIATION RESISTANCE AS A FUNCTION OF CURRENT DISTRIBUTION



R_a = Equivalent Antenna Resistance ($r_a + r_p$)

C_a = Equivalent Antenna Capacity

E_o = Open Circuit Induced Voltage

L_t = Tuning Inductance

R_{L_t} = Loss Resistance Associated With the Tuning Inductance

C_t = Variable Tuning Capacity

R_{C_t} = Loss Resistance Associated With Tuning Capacity

C_b = Input Capacity to Cathode Follower

R_{eq} = Equivalent Noise Resistance of Cathode Follower

R_k = Cathode Resistor

C_c = Coupling Capacity

R_T = Terminating Resistor Equal to Characteristic Impedance of Coaxial Transmission Line

Figure 14. BASE TUNING NETWORK

A GENERALIZED ANALYSIS OF ELECTRONIC ANTENNA BEAM STEERING

By: Dr. P. D. Kennedy, Lockheed Missiles and Space Division

I. INTRODUCTION

The techniques for steering an antenna beam have come to be classified in three categories: mechanical, electromechanical, and electronic. Mechanical steering is exemplified in the traditional search and tracking radars in which the entire antenna assembly is moved about by motor drives. There are also many examples of electromechanical steering, which achieve beam motion by mechanically altering the phase distribution of the feed while maintaining the bulk of the antenna and reflector stationary. Among such examples are the Foster Scanner, the Eagle Scanner, and several systems which ever antedate the microwave era. Finally, electronic beam steering operates entirely without moving parts. This has long been the dream of system engineers who have wanted to control very large antennas at very high scan rates. The terms "large" and "high" are obviously relative, and depend on the drive power and structural strength which are available to combat inertial and other forces. For example, the structure required in an antenna having an aperture of several acres, such as is being proposed to supply motive power to a hovering helicopter, would be prodigous by any standard, while the power available from the solar cells of an interplanetary probe might be barely sufficient for the mechanical drive of a 3-foot communications antenna.

As with all specialists, those working in the electronic scanning field have a tendency to consider only the area which is bounded by the key component of interest to them. Therefore, those things which are common to ferrite scanners, frequency scanners, reactance diode scanners, etc., are usually not apparent. The principal purpose of this report is to give a generalized analysis and to emphasize the common factors of all forms of electronic beam steering. Also, the characteristics of several devices of current interest will be reviewed.

II. CLASSIFICATION OF ELECTRONIC STEERING TECHNIQUES

All electronically scanned antennas of current interest consist of arrays of discrete radiating elements, with control of the beam being accomplished by variation of the phase of the signal at each element. There are two broad classes of phase control devices ("phase shifters") – those in which the signal is passed through some sort of transmission line of appreciable length, and those in which a transmission line is not used and control depends on some lumped parameter element.

The phase shift of a transmission device will be shown to depend on the frequency, length, and phase velocity; this distinction provides the next grouping within this class. Variable-length devices, which entail such factors as mechanical motion of sliding contacts, will not be discussed. However, frequency-controlled steering, as in the Hughes "Frescan" radar and phase velocity-controlled steering, as exemplified by many ferrite systems, will be of principal concern. Since the phase of the output of a frequency converter, or mixer, is governed by the phases of the two inputs, complex superheterodyne systems in which the phase control is performed at various frequency levels are possible. This provides a further basis for technique classification.

Lumped parameter phase shifters usually depend on some form of variable reactance device such as a variable capacitor which might be used to simulate a movable short circuit in a waveguide or a reactance tube used to control the phase of an oscillator. Figure 1 illustrates these principles of system classification.

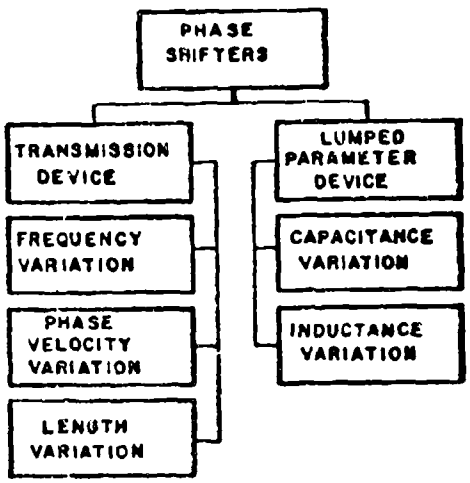


Fig. 1 Phase Shifter Classification (Any of the Above Can Be Applied at Various Frequency Levels)

III. REVIEW OF BASIC FACTORS

When the exponential convention is used to represent an alternating voltage or current, a signal of zero phase angle is expressed as $A\,e^{j\omega t}$, where $\omega = (2\pi)$ (frequency). If this signal is propagated through a medium where the phase velocity is c for a length d, the wavelength in the medium is $\lambda = c/f$ and the output signal is expressed as

$$\begin{aligned} \text{Output} &= B\,e^{j(\omega t - \beta d)} \\ &= B\,e^{j(\omega t - 2\pi d/\lambda)} \\ &= B\,e^{j(\omega t - \omega d/c)} \end{aligned} \qquad (1)$$

In the last form, the term $e^{-j\omega d/c}$ shows the phase lag caused by the passage through the medium.

Antennas are usually analyzed by adding the contributions to the total field at a distant point from the

various radiating elements. In controlling the antenna, the phases of the various signals are adjusted so that these contributions all add in phase at some specified point, and the angle to this point is said to be the direction of the beam. The test point is assumed to be so distant that the rays to it from each element can be assumed to be parallel.

The above observations apply for transmitting antennas, but the results obtained by analysis will also apply to receiving antennas. Conversely, it may sometimes be more convenient to analyze a receiving antenna and apply the results to transmission. Amplitude factors, whether introduced by range considerations, by aperture excitation, or by the effect on the pattern shape of scanning to various angles, will not be considered.

IV. BASIC GENERALIZED ANALYSIS OF ANTENNA SCANNING

1. No Frequency Conversion

Consider two adjacent elements of an array as illustrated by Fig. 2.

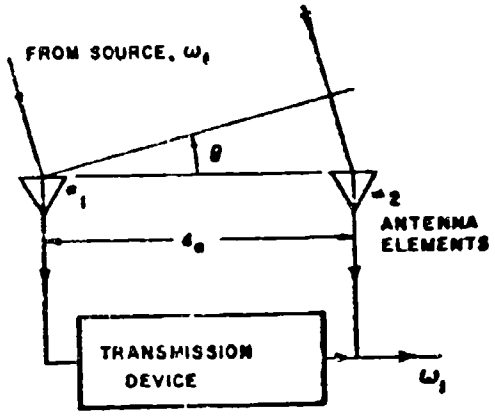


Fig. 2 Array With No Frequency Conversion

At element (2), the phase lag with respect to element (1) is $(\omega_1 d_a \sin\theta)/c_o$ because element (2) is $d_a \sin\theta$ farther from the source. The phase lag introduced by the transmission element is $\omega_1 d_c/c_c$, where the subscript c suggests a length of cable used as a phase shifter. To make these two phases equal for some specified angle so that the "beam is steered to this angle,"

$$\frac{\omega_1}{c_c} d_c - n\,2\pi = \frac{\omega_1}{c_o} d_a \sin\theta^*$$

from which

$$\frac{c_o}{c_c}\frac{d_c}{d_a} - n\,\frac{\lambda_o}{d_a} = \sin\theta$$

where

- c_o = phase velocity in space
- c_c = phase velocity in the transmission device
- d_c = length of transmission device
- d_a = spacing between antenna elements
- n = a positive integer (can be zero)
- λ_o = wavelength in space of signal
- θ = scan angle

and (c_o/c_c) (d_c/d_a) will be called the control variable u.

If n is not zero, two facts are apparent:

- The angle will be dependent on both the signal frequency and the control parameter (c_o/c_c) (d_c/d_a).

  If operation over a band of frequencies is desired, some kind of computer having the frequency as one input will be required to determine the proper control value for a specified angle. Furthermore, the required range of the control parameter will be extended.
- The possibility of multiple lobes exists if several values of n can give real angles. This tends to restrict the bandwidth and scanning angle.

Several curves are given in Fig. 3 to show the effects of varying the control parameter and the signal wavelength for rather small values of (d_c/d_a) (c_o/c_c). Suppose that scanning over ±30° is required and that a 1.4 to 1 frequency range is required as suggested on the figure. If spacing between antenna elements is assumed to be $0.5\lambda_o$ at the shortest wavelength, the control parameter must be variable from 2.5 to 0.95. If the control parameter is larger as illustrated by Fig. 4, the required fractional range of the control variable is reduced.

The possibility of having higher-order beams is always present when the element spacing is greater than about $\lambda_o/2$ and any scanning to positive angles is attempted. That is, for each value of d_a/λ_o, the maximum positive scan angle is limited by the appearance of a lobe at a negative angle. This second lobe may be permitted to exist, under certain circumstances, if it lies outside the range of interest. Figure 3 also illustrates the limitations on scan angle and element spacing for several second-lobe positions.

Equation (2) is perfectly general and applies to all beam steering systems not involving frequency conversions. The type of phase shifter was not specified but only described by its phase velocity and length, either of which may be variable for control-phase velocity as in ferrite devices, ferroelectric devices, traveling wave tubes, etc., or length as in sliding short circuits, traveling helix couplers, trombone line stretchers, etc.

---

* If the phase of each alternate element is reversed, the term $n\,2\pi$ can be π, 3π, 5π, etc., but this is not necessary to the discussion.

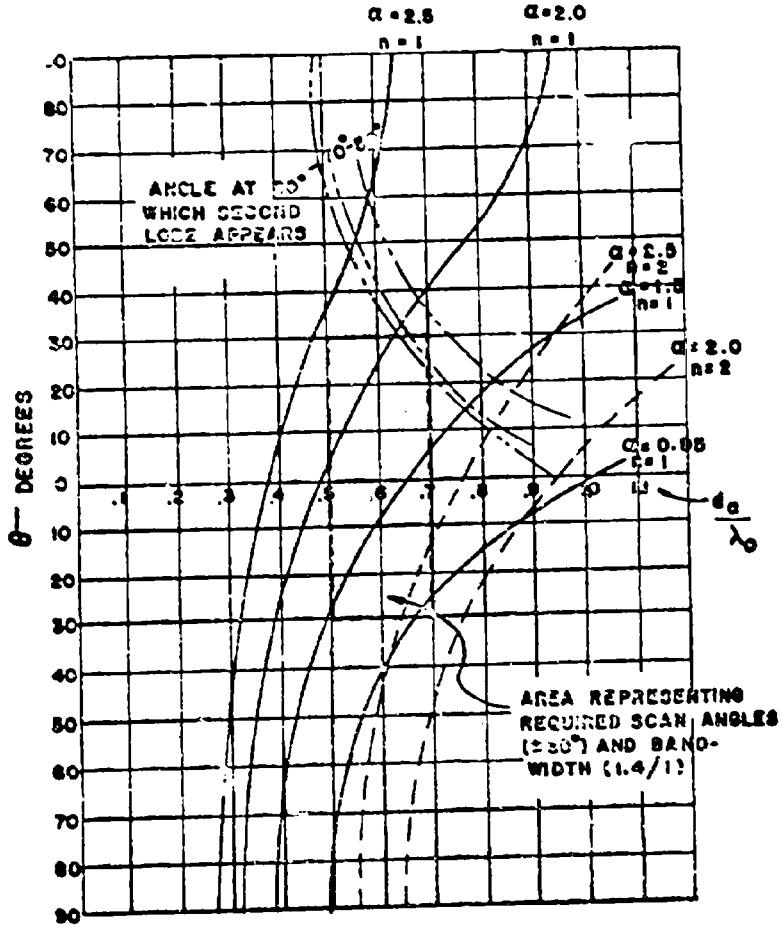


Fig. 3 Scanning System Performance
n = 1,2

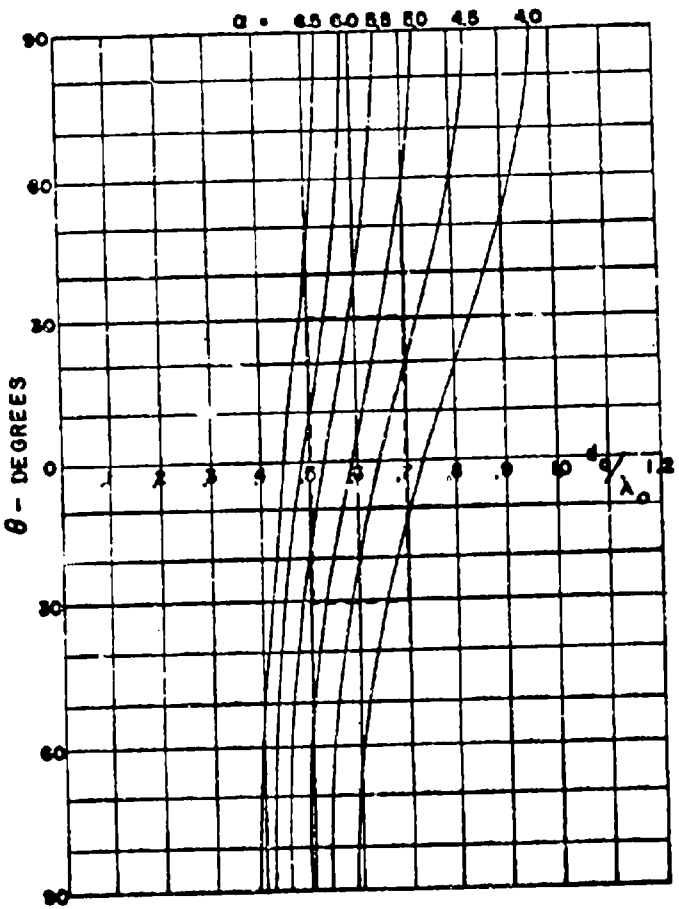


Fig. 4 Scanning System Performance
n = 3

2. Single Frequency Conversion

Assume that the system is arranged as shown in Fig. 5 and that a transmission device is placed in one of the intermediate frequency circuits. The local oscillator voltage has a frequency higher than that of the signal and arrives at each mixer with the same phase. Then, the output of mixer number 2 occurs with a leading phase angle of $(\omega_1 d_a/c_o) \sin\theta$ which is to be compensated by the phase lag of the transmission device. Therefore:

$$\frac{\omega_3}{c_c} d_c - n\,2\pi = \frac{\omega_1}{c_o} d_a \sin\theta$$

$$\frac{\omega_3}{\omega_1} \frac{c_o}{c_c} \frac{d_c}{d_a} - n\,\frac{\lambda_o}{d_a} = \sin\theta \qquad (3)$$

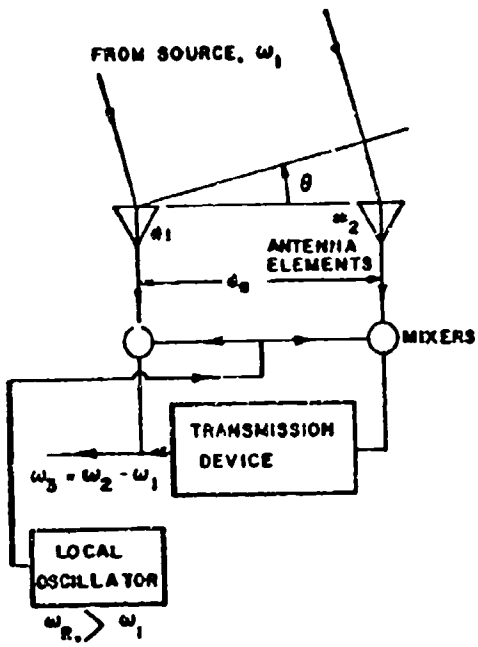


Fig. 5 Array With Single Frequency Conversion

The range of control parameter and scanning limits is exactly as shown in the previous section. However, ω_3 as well as c_c and d_c can be now used to vary the control parameter. For microwave frequencies, ω_3/ω_1 will be small and a large value of $(c_o/c_c)(d_c/d_a)$ will be necessary, but this can easily be done by making the transmission device a length of delay line cable; for example, in RG-63 A/U, c_o/c_c is about 40. Furthermore, an advantage is gained from the presence of ω_1 in the control parameter, because it acts to reduce the required range of ω_3 to cover given scanning angles.

The preceding discussion does not imply that a system involving a frequency conversion is useful only when the intermediate frequency is changed to vary the control parameter. On the contrary, the control may be obtained by means of changes in the phase velocity or length of the I-F delay line (Ref. 1); and if this can be done more conveniently than by changing the corresponding characteristics of the signal frequency transmission lines, justification for the frequency conversion is provided.

3. Double Frequency Conversion

Assume now that the system is arranged as shown in Fig. 6 and that a transmission device is placed in the circuit of the second local oscillator.

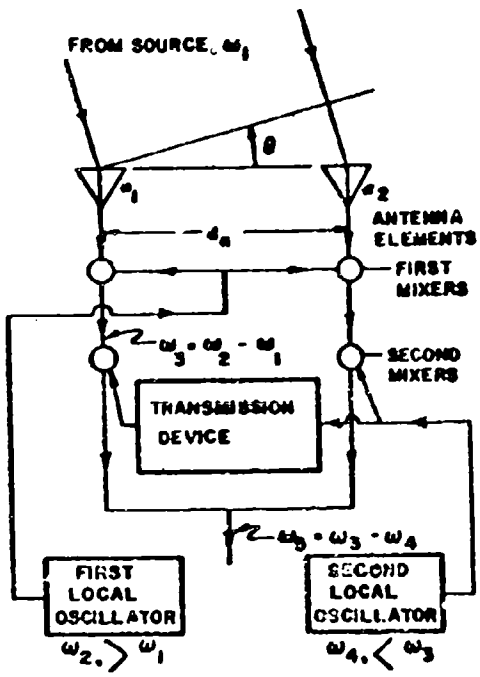


Fig. 6 Array With Double Frequency Conversion

Accounting for the various phase shifts leads to the following equation for the condition when the contributions to the output are in phase:

$$\frac{\omega_4 d_4}{c_4} - n\,2\pi = \frac{\omega_1 d_a}{c_o} \sin\theta$$

$$\frac{\omega_4}{\omega_1} \frac{c_o}{c_4} \frac{d_4}{d_a} - n\,\frac{\lambda_o}{d_a} = \sin\theta \qquad (4)$$

where c_4 and d_4 are the phase velocity and length of the transmission element in the second local oscillator circuit. This equation is similar to Eq. (3) inasmuch as the control parameter can be readily identified and

it happens to contain an adjustable frequency term. This system has the interesting property of having independent controls for frequency tuning (which is varied by the first oscillator, as in any superheterodyne system) and for angular scanning (which is varied by the second oscillator). Actually, the phase shifts could be introduced at any of five frequency levels (and a similar multiplicity of opportunities for placement of the phase shifter in the single conversion system), but the arrangement shown seems to be the most interesting. As a matter of fact, it might be possible to use some sort of lumped parameter elements to obtain the second oscillator phase shifts at some convenient low and fixed frequency, which would justify the double conversion. Furthermore, it may be noted that the phase shift element in Fig. 6 is not required to pass the modulated signal, and this may be an advantage where the bandwidth of the device is limited.

V. PRACTICAL PHASE SHIFT DEVICES

1. General Requirements

It has been shown that analysis of any beam steering system leads to the equation

$$\sin \theta = a - n \frac{\lambda_0}{d_a}$$

where the control parameter a only has the form

$$a = \frac{\omega_n}{\omega_1} \frac{c_0}{c_c} \frac{d_c}{d_a}$$

To show how this equation is applied in practice will entail the investigation of details about devices whose frequency, phase velocity, or length can be controlled. (Length will be excluded from this discussion of electronic techniques.)

Assuming that such devices can be found, by what criteria should they be judged? It seems most pertinent to consider the following points of interest:

- Reliability in their environment
- Weight and size
- Power consumption
- Loss, gain, and noise
- Ease of control
- Applicability in various frequency bands
- Present state of development

The following sections will discuss how several devices of current interest meet the above requirements.

2. Techniques of Phase Velocity Control

In an unbounded medium of permittivity $\epsilon_r \epsilon_0$ and permeability $\mu_r \mu_0$, the phase velocity is

$$c = \frac{1}{\sqrt{\epsilon_r \mu_r} \sqrt{\epsilon_0 \mu_0}} = \frac{c_0}{\sqrt{\epsilon_r \mu_r}}$$

In general, then, one seeks materials for use in phase shifters whose relative permittivity and permeability are susceptible of electrical control. The following immediately come to mind:

- Ferroelectric materials (variable ϵ_r)
- Gas discharge tubes (variable ϵ_r)
- Ferrites (variable μ_r)

Next to be considered are "artificial" transmission lines composed of lumped capacitance and inductance. The following devices, which are equipped for electrical control, are available:

- Capacitor diodes (variable c)
- Saturable reactors (variable L)

Furthermore, the velocity of a traveling wave which interacts with an electron beam can be varied by changing the characteristics of the beam, so another useful phase shifter is the

- Traveling-wave tube

In a bounded medium where the propagation is not purely TEM, the phase velocity depends on the ratio of signal wavelength to cutoff wavelength in the medium. For example, in hollow waveguides,

$$c = \frac{c_0}{\sqrt{\epsilon_r \mu_r} \sqrt{1 - (\lambda_0/\lambda_{co})^2}}$$

Devices which vary the cutoff wavelength mechanically will not be considered, but it seems most appropriate to classify a dispersive waveguide as an element whose phase shift depends, in part, on the change of phase velocity with frequency.

3. Phase Shift Devices

a. Ferroelectrics

The term "ferroelectric" (Refs. 2 and 3) is applied to dielectric materials which exhibit a nonlinear curve relating D and E (similar to the B-H curves of ferromagnetic materials) and for which, as a consequence, the dielectric constant can be controlled by varying the electric field strength. This suggests that a two-conductor transmission line, for example, might be imbedded in such material and the velocity of propagation might be controlled by varying a d-c bias between the conductors. Unfortunately, the cause of the nonlinear behavior -- a distortion of the crystal structure -- also makes these materials sensitive to temperature and mechanical stress. Furthermore, the materials would be considered lossy in comparison with the usual dielectrics, since the loss tangent is in the range of 0.05 to 0.15 for most ceramic ferroelectrics which might be used. For example, a loss tangent of 0.1 means that a transmission line with an electrical length of two wavelengths would have an attenuation of about 5 db.

Another disadvantage is that rather high voltages would be required - perhaps several thousand volts in small coaxial cables or several hundred volts in stripline of a few mils separation. The ferroelectrics are good insulators, but the generation and control of such voltages would surely present problems.

It appears, then, that the ferroelectrics hold some promise for the future, and interest in their

microwave properties is increasing, but they do not seem suitable for immediate application in antenna systems.

b. Gas Discharge Tubes

A derivation of the electrical properties of a gas in which free electrons are present shows that the relative permittivity is less than unity and is related to the electron density (Ref. 4). This has suggested that a phase shifter be constructed of a microwave transmission line in which the gaseous dielectric can be ionized to a controllable degree by an auxiliary electrical discharge. This principle has been demonstrated in a relatively crude way (Ref. 5), but such a device has many disadvantages. The attenuation caused by the energy loss associated with collisions between electrons and larger particles may be excessive; transmitting systems are handicapped by the limited power-handling ability of the low-pressure gas; and receiving systems are handicapped by the appreciable noise generated by the discharge. However, the possibility of using gas discharge phase shifters is intriguing and may eventually become feasible.

c. Ferrites

It is gross simplication to say that a ferrite material has a variable μ ; actually the magnetic moment of the spinning electron causes the material to react differently to the two senses of circular polarization, and this effect happens to be sensitive to the strength of the applied d-c magnetization (Refs. 6 and 7). Therefore, ferrite devices must be explained on the basis of this anisotropy and in terms of the existence of circularly polarized components of the r-f field. Generally speaking, most of the waveguide configurations require rather high field strengths in the magnetic fields used for control, so that relatively high currents are required in the coils. Where iron is used in the magnetic circuit, the devices suffer hysteresis in the control current-phase shift curve. Consequently, the system designer faces serious problems in the design of the control apparatus, not only because of the hysteresis but also because of the temperature sensitivity of the materials, and there is little hope for the feasibility of really lightweight and compact equipment having low power consumption (these adjectives being entirely relative, of course).

However, at least one type of ferrite phase shifter has appeared which is remarkably economical of control power – the so-called Reggia-Spencer phase shifter. This consists of a ferrite rod centered in a rectangular waveguide and a solenoid to produce a longitudinal control field. The inventors theorized that the large amounts of phase change occurred because the polarization rotation of the field in the rod is suppressed by one waveguide dimension being beyond cutoff (Ref. 8). Others have proposed that the applied magnetic field tends to concentrate more energy in the rod, where the high dielectric constant causes considerable reduction of the phase velocity (Refs. 9, 10, and 11). In any case, this type of device seems the most promising in terms of ease of control, and control power requirements of 1/3 to 1/2 watt per phase shifter have been obtained.

Most of the developments reported in the literature have been made in waveguide devices, and the majority have been demonstrated at X-band (Ref. 12), although some stripline and coaxial components have been announced. The ferrite materials themselves have been improved so as to be suitable for application low in the UHF band; for example, some data were recently presented on a microstrip phase shifter at 400 and 500 megacycles (Ref. 13).

d. Capacitor Diodes

It is well known that a sliding short circuit can be used as a component in a phase shifter, for example, as in the E-H tuner. The variable reactance of a diode capacitor makes possible an analogous device in which the voltage-controlled capacitance can be used to produce the same result as the sliding short. The varactor diodes being manufactured for use in parametric amplifiers are of such high Q that the loss on reflection will be about 1/2 db and have a sufficient range of capacity to produce a phase shift of about one-quarter wavelength (Refs. 14 and 15). However, the relation between bias and phase shift is quite nonlinear, and the power-handling ability is limited.

The use of capacitor diodes as phase shifting elements is, nevertheless, a technique with very great potential, particularly for light, compact equipment. One advantage is that the control power requirements are very low, since the diode is biased in the reverse direction. Others are that the diode is very small, should have unlimited life, and is relatively insensitive to temperature variations. One might visualize a rather complex stripline assembly incorporating diodes which are permanently soldered in place and controlled by printed circuitry which parallels the r-f lines. (A coax version is described in Ref. 16.) Further improvement in diode performance may make this concept feasible.

At frequencies up to about 20 megacycles, the low-frequency type of silicon capacitor can be used as a variable element in an artificial transmission line. Such capacitors show approximately a 4/1 capacitance change for a bias range from zero to about 20 volts. If the inductance of the circuit is varied also, the phase shift may be obtained while keeping the characteristic impedance constant, thereby improving the efficiency of the circuit.

e. Saturable Reactors

Variable inductors are commercially available for the frequency range where an artificial transmission line circuit might be used. Such units are capable of approximately a 4/1 inductance change, with about 1 watt of control power required at the extreme inductance value.

f. Traveling Wave Tubes

Traveling wave tube amplifiers have many applications in the microwave art because of their relatively wide bandwidth and the ease of construction and use. Because the phase velocity of the wave progressing along the helix is sensitive to the helix d-c potential, these tubes seem destined to play an important part in the future of electronically scanned antennas as well, particularly where their power output or gain can be used in a transmitting or receiving application, respectively.

Traveling wave tubes have been built for all frequencies within the UHF and microwave bands, and a given design will usually cover an octave bandwidth. They are typically about 1 inch in diameter and 15 inches long and weigh 1 or 2 pounds, exclusive of the focusing magnet. The power requirement of a low-power-receiving type of tube is about 10 watts, exclusive of magnet power supply. The disadvantage which accompanies these desirable features is the focusing solenoid, which adds perhaps 10 pounds and 100 watts to the weight and power requirements. Presumably, all traveling wave tubes will ultimately be focused by either very light permanent magnets or electrostatic structures.

VI. CONCLUSIONS

It has been shown that a great variety of electronic antenna scanning systems can be represented by one equation in which a control parameter can be identified. Adjustment of this parameter can be obtained by variation of frequency, phase velocity, or length, and a discussion of several phase shift devices illustrated how they can be considered from this point of view. The use of the analysis presented particularly shows the requirements on the control parameter to obtain scanning over a band of signal frequencies.

REFERENCES

1. J. R. Copeland, "Electronic Scanning: A Memorandum on Heterodyne Conversion," Report 903-7, Antenna Laboratory, The Ohio State University Research Foundation, 15 Nov. 1950.

2. W. J. Gemulla and R. D. Hall, "Ferroelectrics at Microwave Frequencies," The Microwave Journal, Vol. 3, No. 2, Feb. 1960 pp. 47 - 51.

3. L. Davis and L. G. Rubin, "Some Dielectric Properties of Barium-Strontium Titanate Ceramics at 3000 Megacycles," Journal of Applied Physics, Vol. 24, Sep. 1953, pp. 1194 - 1197.

4. L. D. Smullin and P. Chorney, "Properties of Ion Filled Waveguides," Proceedings of the Institute of Radio Engineers, Vol. 46, Jan. 1958, p. 360.

5. R. C. Jones and G. R. Lowrey, Jr., "Plasma Resonance Scanners," The Microwave Journal, Vol. 2, No. 4, Apr. 1959, pp. 21 - 28.

6. K. J. Button, "Historical Sketch of Ferrites and Their Microwave Applications," The Microwave Journal, Vol. 3, No. 3, Mar. 1960, pp. 73 - 82.

7. B. Lax, "The Status of Microwave Applications of Ferrites and Semiconductors," Transactions of the Institute of Radio Engineers, Vol. MIT-6, Jan. 1958, pp. 5 - 18.

8. F. Reggia and E. G. Spencer, "A New Technique in Ferrite Phase Shifting for Beam Scanning of Microwave Antennas," Proceedings of the Institute of Radio Engineers, Vol. 45, Nov. 1957, pp. 1511 - 1517.

9. A. Clavin, "Reciprocal Ferrite Phase Shifters in Rectangular Waveguide," Transactions of the Institute of Radio Engineers, Vol. MIT-8, Jul. 1958, p. 334.

10. P. A. Rizzi and B. Gallia, "Rectangular Guide Ferrite Phase Shifters Employing Longitudinal Magnetic Fields," Proceedings of the Institute of Radio Engineers, Vol. 47, Mar. 1959, p. 446.

11. J. A. Weiss, "A Phenomenological Theory of the Reggia-Spencer Phase Shifter," Proceedings of the Institute of Radio Engineers, Vol. 47, Jun. 1959, pp. 1130 - 1137.

12. F. E. Goodwin and H. R. Seaf, "Volumetric Scanning of a Radar With Ferrite Phase Shifters," Proceedings of the Institute of Radio Engineers, Vol. 47, Mar. 1959, p. 453.

13. C. M. Johnson, "Bandwidth of Ferrite Phase Shifters for Phased Array and Direction Finding Use," Proceedings of the Institute of Radio Engineers, Vol. 47, Sep. 1959, p. 1668.

14. A. Uhlir, Jr., "The Potential of Semiconductor Diodes in High Frequency Communications," Proceedings of the Institute of Radio Engineers, Vol. 46, Jun. 1958, pp. 1099 - 1115.

15. N. Houlding, "Measurement of Varactor Quality," The Microwave Journal, Vol. 3, No. 1, Jan. 1960, pp. 40 - 45.

16. W. G. Swarner, "A Voltage Controlled Radio Frequency Phase Shifter," Report 903-5, Antenna Laboratory, The Ohio State University Research Foundation, 1 Aug 1958.

THE EFFECT OF FIELD WIRE STABILITY ON THE MAXIMUM LENGTH OF LOOP

M.N. Feldman and G.P. Tripp
USASRDL, Fort Monmouth, N.J.

Summary. A study reveals that field wire instability will impose severe limitations on system design and performance. In this paper a general equation is derived for maximum loop length in terms of stability for given loop loss. Charts have been prepared showing the maximum usable loop length in miles for wet and dry Wire WD-1 with 3 mi. repeater spacing. It is established in the paper that the absolute maximum length of loop is dependent on the following factors: maximum allowable loop loss, wire stability, wire attenuation and unattended repeater spacing. The maximum allowable loop loss is fixed by system considerations. The wire stability and attenuation are design factors under the control of the cable engineer. The unattended repeater spacing is the compromise made by the equipment design engineer. The smaller the spacing the closer the approach to the absolute maximum loop length and the greater the number of repeaters.

INTRODUCTION

It is generally accepted that there exists a maximum value for the usable length of repeatered transmission line. Ordinarily, this is based on the total noise picked up by the conductors and amplified by the repeaters. The maximum value is reached when the signal-to-noise ratio reaches the prescribed maximum value. Thus, even though the signal strength may yet be restored, the high level of incident noise has impaired the man or machine intelligibility of the transmitted signal.

In the case of military field wire, another factor, more powerful in effect than the noise, plays a length limiting role. This factor is attenuation stability. Attenuation stability (to be called stability in this paper) is defined as the difference in attenuation between a mile of dry wire and a mile of wet wire. This results from the increase in the capacitance and conductance parameters over the dry condition when the wire pair becomes wet.

In this paper we will derive a general equation for maximum repeatered line length. We will show that, given the maximum allowable loss, the maximum line length is a function of the following factors: stability, unit attenuation, and repeater spacing.

THE PROBLEM

The sophisticated consideration of the line loss has a tremendous impact on its communication system. It can be shown that the anticipated line loss is the controlling factor in allocating allowable noise level per trunk. A 3 db decrease in anticipated maximum line loss would allow a 3 db increase in allowable trunk noise level. This, in turn, would reduce the transmitter power requirements by almost 50% in a radio trunk. Such reduction in transmitter power requirements would result in significant reductions in operating cost, size, and weight, of the entire communication system and its components.

In the light of the preceding discussion our purposes in writing this paper become apparent.

(1) To develop and exhibit the various cable characteristics which contribute to transmission line performance.

(2) To develop the relationship between the various cable characteristics and the transmission line performance.

(3) To punctuate the importance of engineering desirable characteristics into transmission line cable.

DEVELOPING THE EQUATION

The requirements imposed on a communication system for military use include: easy installation, foolproof administration, simple maintenance and low cost. The wire subscriber loop especially must meet these requirements. For this reason it was decided to use fixed gain type of repeaters and amplifiers. The use of these repeaters would be prescribed at the end of each section that exceeded a predetermined length in miles. The schematic of Figure 1 illustrates a general subscriber loop that is L miles long and the prescribed repeater spacing is 3 miles.

The relationships that will be developed are general. To punctuate the impact of our development, an actual set of conditions will be used. The conditions will consist of the use of Field Wire, WD-1 and the application of the 9 db maximum loop loss specification.

The equations will be developed using symbols having interpretations as follows:

L = Total length of loop, miles

L_{usable} = Total usable length of loop, miles

L_{last} = Length of loop to last repeater, miles

3 = Repeater spacing, miles

X = Maximum allowable loss per loop, db

X_w = Attenuation per mile of wet wire, db

X_D = Attenuation per mile of dry wire, db

C = Stability Constant of the wire, $\frac{(X_w - X_D)}{X_w}$

A maxim to be observed is: the gain of the repeaters should not exceed the losses of the line. Failure to observe this maxim could result in: (1) high signal level at the listener's ear; (2) circuit singing; (3) overloaded transmission circuit elements with resultant cross-talk, distortion, etc. Hence, with unstable cable, the fixed repeater gain must be keyed to the attenuation of the cable when dry, its low loss condition. The loop loss, for dry cable, would never be greater than, the loss of one repeater section and never less than zero. This is illustrated in the transmission loss diagram of Figure 2. The lines AB, CD, EF, etc. represent the insertion gain of the repeaters and the lines OA, BC, DE, etc. represent the cable losses.

When the cable changed to wet, the gain of the repeaters would be less than the cable loss. The difference in loss between that of the wet cable and that of the dry cable must be borne by the system. There is no simple and effective way of compensating for it. If repeaters could be applied on a distributed basis (S=0), there would be no "left-over" effect and maximum usable loop length would be:

$$L_{usable} = \frac{X}{X_w - X_D} = \frac{X}{CI_w} \qquad (1)$$

The diagram of Figure 3 shows the transmission loss along the wire loop when the cable is wet. The transmission loss curve for repeatered dry WD-1 has been modulated by a function that is equivalent to the difference in loss between the wet cable and dry cable. Under these conditions the loop loss soon exceeds the 9 db limit specified. The "left-over" effect is illustrated in Figure 3 at the 5.5 mile point. At this point, the transmission loss starts to exceed the specified maximum. This is an unacceptable condition even though the next repeater will restore the line to specification quality. This effect imposes another limitation on the maximum length of line.

The Transmission Loss Diagram of Figure 4 emphasizes the last repeater section. The maximum usable length of line is fixed by the point where the transmission loss crosses the line designating the "Spec Limit for Loop Loss". In this figure this occurs at "F". Then, the last repeater, would appear at "=D". When the gain of this repeater is exactly equal to the distance between the L ($X_w - X_D$) line and the limit line, we have the conditions for maximum loop length. It can be shown that the usable loop length represented by B is:

$$L_{usable} = \frac{X}{CI_w} - S\left(\frac{(1-C)^2}{C}\right) \qquad (2)$$

The diagram of Figure 4 shows the location of the last repeater that would give the maximum usable loop length for a given repeater gain. This repeater gain is considered in terms of miles of dry field wire loss and designated by the letters "SX_D". The spacing must be so selected that the quotient, of the length of line to the last repeater divided by "S", be a whole number. It can be shown that the length of line to the last repeater, AD is:

$$L_{last} = \frac{X}{CI_w} - S\left(\frac{1-C}{C}\right) \qquad (3)$$

Then, the maximum number of repeatered sections would be:

$$\frac{L_{last}}{S} = \frac{X}{SCI_w} - \left(\frac{1-C}{C}\right) \qquad (4)$$

APPLICATION OF THE EQUATIONS

Now that the equations have been developed let us see what they can tell us. The first equation, Equation (1), tells us that, even with distributed parameter type repeaters, the maximum length of acceptable loop depends on the cable unit loss and the stability. The next equation, Equation (2), includes the effect of the repeater spacing when it is greater than zero.

To further illustrate the theme of this paper, the equations were applied to the performance of three types of cable. The characteristics of the considered cables were chosen as follows:

(1) Those of present Field wire, WD-1
(2) Those of a cable having same attenuation as WD-1 but lower stability constant.
(3) Those of a cable having higher attenuation than WD-1 and lower stability constant.

The characteristics of the above cables were cranked into the equations. The contents of Table I tabulate the cable data and the maximum loop length figures calculated for various assumed repeater spacings. These spacings were selected to give an exact integer for the number of repeatered sections. The first computed maximum loop length in each group of data is for the condition where the repeater spacing is zero and the number of repeaters is infinite. This is an impractical design, but illustrates the limiting case if distributed parameter repeaters could be made. Continuing down the column, in these groups, it is seen that as the number of repeaters reduces or the repeater spacing increases to practical values, the maximum loop length shrinks drastically.

CONCLUSIONS

Our dissertation has established that the absolute maximum length of loop is dependent on the following factors: maximum allowable loop loss, wire stability, wire attenuation, and unattended repeater spacing. The maximum allowable loop loss is established by system considerations, the wire stability and wire attenuation are design factors under the control of the cable designer. The unattended repeater spacing is the compromise made by the equipment

designing engineer. The smaller the spacing the closer the approach to the absolute maximum loop length and the greater the number of repeaters required.

The use of wire having stability and attenuation characteristics of WD-1 imposes an absolute maximum loop length limit of 9.75 miles and a practical maximum loop length of 6.9 miles, with 2.6 mile repeater spacing. This same wire with its stability improved to 10% will have vastly improved performance. The loop length asymptote, is improved by a factor of approximately four. Also, the maximum practical loop length, with 2.7 mile repeater spacing, is improved by a factor of about two. These are the things that are obtained by improving the stability.

The primary purpose of this paper has been to divulge the "Effect of Field Wire Stability on the Maximum Length of Loop". Another purpose was to emphasize the importance of this characteristic. It is the hope of the authors that, in the future, when military communication system planners consider the loop, the cable stability be remembered.

ACKNOWLEDGEMENTS

The authors wish to acknowledge with gratitude the interest and encouragement displayed, by Mr. G. W. Bartle, in the subject matter covered by this paper. Subject to additional gratitude is the patience and cooperation exercised by our secretary Mrs. D. M. Neilson through whose efforts all of our deadlines were met.

APPENDIX

Proof of Last Section Equations.

Consider Figure No. 4.

AC is length of loop between last repeater and end of ideal line.

AB is length of loop between last repeater and end of practical line.

BC is (AC - AB). This is the difference in length between the ideal and practical line.

The triangle formed by ACF is similar to the triangle formed by OIF hence

$$(11)\quad \frac{AC}{CF} = \frac{1}{X_w - X_D}$$

Since CF = AD = SX_D and

$$(12)\quad AC = \frac{SX_D}{X_w - X_D} \qquad X_D = X_w - CX_w \text{ by definition then}$$

$$AC = \frac{S\,(X_w - CX_w)}{CX_w}$$

$$(13)\quad AC = S\left(\frac{1-C}{C}\right)$$ This is the correction term applied to equation (1) in report to give equation (3)

The triangle formed by ABD is similar to the triangle formed by OIX hence:

$$(14)\quad \frac{AB}{AD} = \frac{1}{X_w}$$

$$(15)\quad AB = \frac{AD}{X_w}$$

$$= \frac{SX_D}{X_w}$$

$$= \frac{S(X_w - CX_w)}{X_w}$$

$$(16)\quad AB = S\,(1-C)$$

$$(17)\quad BC = AC - AB$$
$$BC = \left(S\left(\frac{1-C}{C}\right) - S\,(1-C)\right)$$

$$(18)\quad BC = \left[\frac{1-C}{C} - (1-C)\right] S$$

$$(19)\quad BC = \left[\frac{(1-2C+C^2)}{C}\right] S$$

$$(20)\quad BC = S\left(\frac{1-C}{C}\right)^2$$ This is the correction applied to equation (1) in the report to give equation (2) in the report

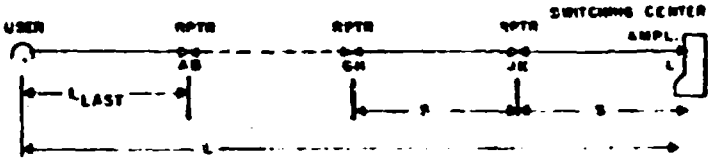


FIG. 1 — SCHEMATIC OF WIRE SUBSCRIBER LOOP

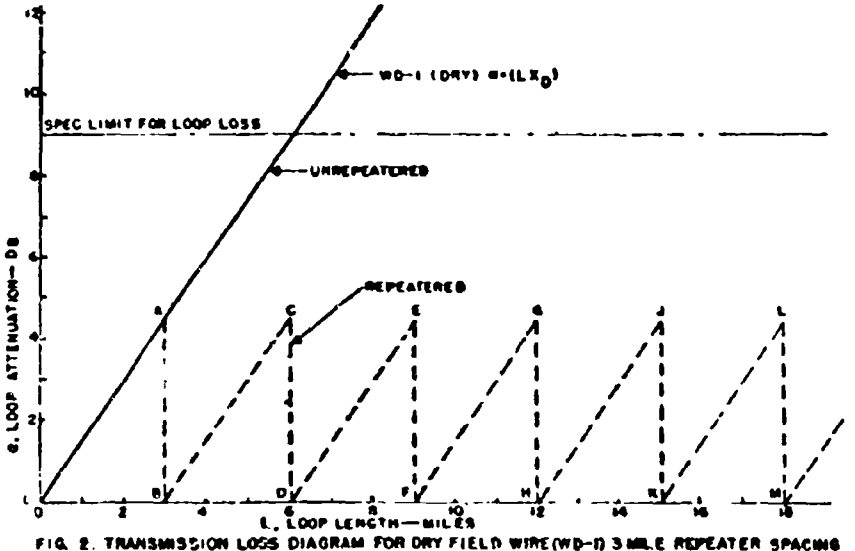


FIG. 2. TRANSMISSION LOSS DIAGRAM FOR DRY FIELD WIRE (WD-1) 3 MILE REPEATER SPACING (AT 1kc)

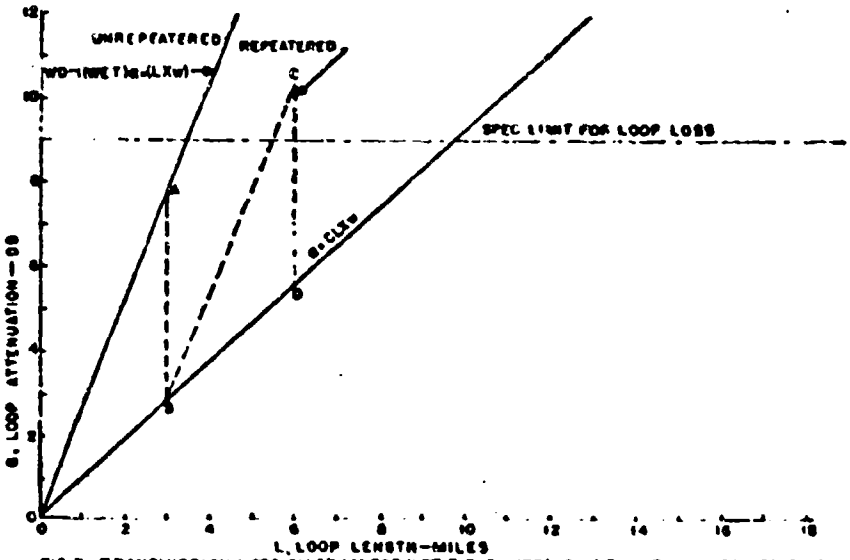


FIG. 3—TRANSMISSION LOSS DIAGRAM FOR WET FIELD WIRE (WD-1) 3 MILE REPEATER SPACING (AT 18c).

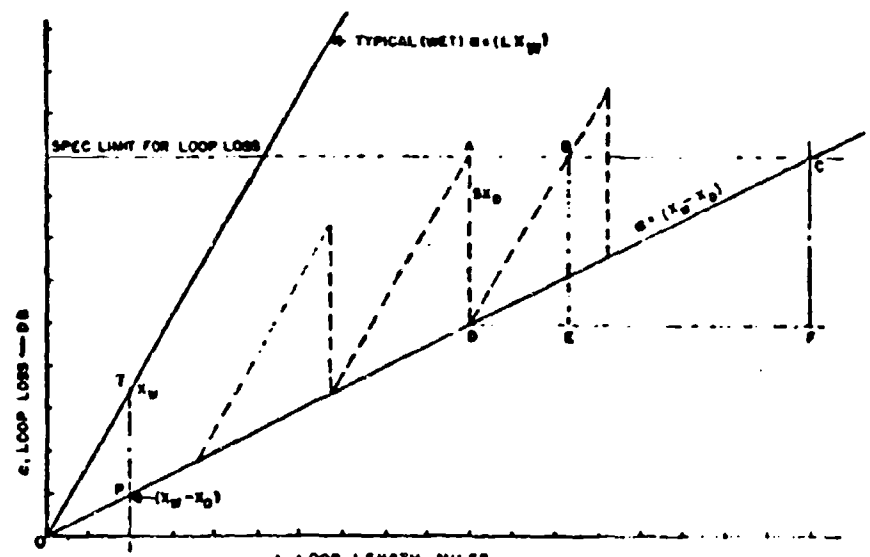


FIG. 4—TRANSMISSION LOSS DIAGRAM EMPHASIZING LAST REPEATER SECTION

WIRE TYPE	ATTENUATION α_w (dB/MILE)	STABILITY C	SPACING S (MILES)	NO. OF REPEATERS	MAX. LOOP LENGTH (MILES)
[illegible]	3.6	[illegible]	[illegible]	[illegible]	25.6
			[illegible]	[illegible]	[illegible]
			[illegible]	[illegible]	[illegible]
			[illegible]	[illegible]	[illegible]
			2.27	2	[illegible]
[illegible]	2.93	[illegible]	[illegible]	[illegible]	9.75
			[illegible]	[illegible]	[illegible]
			[illegible]	[illegible]	[illegible]
			[illegible]	[illegible]	[illegible]
			[illegible]	[illegible]	[illegible]
			[illegible]	[illegible]	[illegible]
COMPOSITE	2.53	[illegible]	[illegible]	[illegible]	[illegible]
			[illegible]	[illegible]	[illegible]
			[illegible]	[illegible]	[illegible]
			[illegible]	[illegible]	[illegible]
			2.38	[illegible]	[illegible]
			2.73	[illegible]	[illegible]
			[illegible]	[illegible]	[illegible]

RING ARRAYS

By: Mr. W. D. Nelson, Heavy Military Electronics Department, General Electric Co.

Ring arrays, as presently pictured, consist of several concentric rings, each made up of a number of elements. An artist's conception of the array is shown in Figure 1.

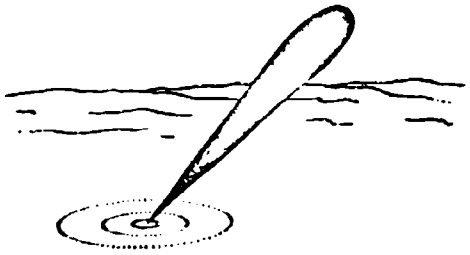


FIGURE 1

For most applications the array would probably be placed flat on the earth's surface with no vertical extent except that of the elements. There are however applications such as telemetering in which the elements of the array are directive and more en masse. Ring arrays can be built with no vertical projection at all, if their coverage is restricted to higher elevation angles. The use of slots for elements in the array is possible. Vertical polarization is necessary in this array for the very low angle coverage, but at the higher elevation angles any polarization can be used.

By their nature, flat arrays must have a very large horizontal dimension for a narrow beamwidth near the horizon. Any attempt to reduce the horizontal size involves either adding a vertical dimension or mechanical movement of directive elements, such as end-fire radiators or reflectors.

Due to the array's large horizontal extent, the beam of a flat array becomes very narrow near broadside. This narrowing can be eliminated by dropping the outside rings when the beam is directed to near-zenith positions or by properly phasing the rings with respect to each other.

History

The principle of feeding antennas in a ring with currents of the same numerical value but with a phase that increases uniformly along the circle was investigated as early as 1936 by H. Chireix.[1] An antenna similar to the Chireix type, and one involving concentric rings was proposed and studied by Hansen and Woodyard[2] in 1938 and further investigated by Hansen and Hollingworth in 1939.[3] Ring arrays with axial, tangential, or radial elements were explored by Knudsen as recently as 1956.[4] A two ring steerable beam communications antenna was built in 1953[5] at Wright Field, Dayton, Ohio. The single ring has been investigated by several people (6)(7)(8)(9) at General Electric.

Pattern Equations

In calculating the far field pattern of a ring array, the coordinate system shown below is used.

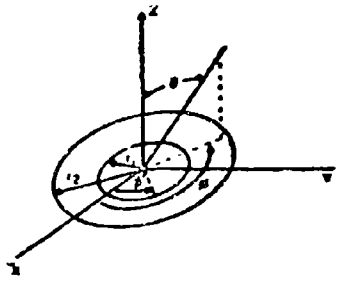


FIGURE 2

A general array consists of N rings and many elements per ring. These antenna elements are assumed omnidirectional. First let us consider a single ring with radius r_m. The center of the ring is used as a phase reference. The array factor is easily found by adding the contributions of each element to the field at some general far field point P.

$$F_m(\theta,\phi) = \sum_{\ell=1}^{B} I_\ell \exp j\left[\alpha_\ell - \frac{2\pi r_m}{\lambda} \sin\theta\cos(\phi-\phi'_\ell)\right] \qquad (1)$$

Where I_ℓ = amplitude of illumination of element ℓ

α_ℓ = phase of illumination of element ℓ

ϕ'_ℓ = location on the ring of element ℓ.

If the phase of the individual elements in the ring is adjusted to cancel the phase in equation (1) at an angle θ_o, ϕ_o the beam formed by the array will be steered to the direction θ_o, ϕ_o. The illumination of the ring of radius r_m is

$$I = \sum_{\ell=1}^{B} I_\ell \exp j\left\{\left[\frac{2\pi r_m}{\lambda} \sin\theta_{om}\cos(\phi_{om}-\phi'_\ell)\right]\right\} \qquad (2)$$

and the far field pattern is

$$F_m(\theta,\phi) = \sum_{\ell=1}^{B} I_\ell \exp j\left\{\left[\frac{2\pi r_m}{\lambda}\left[\sin\theta_{om}\cos(\phi_{om}-\phi'_\ell) - \sin\theta\cos(\phi-\phi'_\ell)\right]\right]\right\} \qquad (3)$$

If the number of elements in the ring becomes large the summation becomes an integral.

$$g_m(\theta,\phi) = \int_0^{2\pi} r_m A_m(\phi') \exp j\left\{\frac{2\pi r_m}{\lambda}\right.$$

$$\left. \times \left[\sin\theta_{om}\cos(\phi_{om}-\phi') - \sin\theta\cos(\phi-\phi')\right]\right\} d\phi \qquad (4)$$

In the general case there will be M rings, and the far field array pattern can be found by summing the effects of the individual rings.

$$g(\theta,\phi) = \sum_{m=1}^{M} g_m(\theta,\phi)$$

$$= \sum_{m=1}^{M} \int_0^{2\pi} r_m A_m(\phi') \exp j\left\{\frac{2\pi r_m}{\lambda}\left[\sin\theta_{om}\cos(\phi_{om}-\phi')\right.\right.$$

$$\left.\left. - \sin\theta\cos(\phi-\phi')\right]\right\} d\phi' \qquad (5)$$

Following Knudsen[4] the trigonometric substitutions below will be made.

$$\cos_m\beta = \frac{\sin\theta\cos\phi - \sin\theta_{om}\cos\theta_{om}}{\rho_m}$$

$$\rho_m = r_m\left[(\sin\theta\cos\phi - \sin\theta_{om}\cos\phi_{om})^2\right.$$

$$\left. + (\sin\theta\sin\phi - \sin\theta_{om}\sin\phi_{om})^2\right]^{\frac{1}{2}}$$

With these substitutions the far field pattern can be written as

$$g(\theta,\phi) = \sum_{m=1}^{M} \int_0^{2\pi} r_m A_m(\phi') \exp -j\left[\frac{2\pi}{\lambda}\rho_m\cos_m(\beta-\phi')\right] d\phi' \qquad (6)$$

If $A_m(\phi)$ is a constant in ϕ' the integration can be performed.

$$g(\theta,\phi) = \sum_{m=1}^{M} 2\pi A_m r_m J_o\left(\frac{2\pi\rho_m}{\lambda}\right) \qquad (7)$$

Let us briefly investigate a single ring of radius a. The normalised array pattern is

$$g_1(\theta,\phi) = J_o\left(\frac{2\pi\rho_a}{\lambda}\right) \qquad (8)$$

The far field pattern of a single ring antenna is thus a zero order Bessel function. The first side lobe of such a pattern is only 8db down from the peak. The 3db points occur at $\frac{2\pi\rho}{\lambda}A = 1.127$. These relationships hold independent of the form of ρ_a. The beam shape as a function of θ and ϕ changes, of course, depending on the value of θ_o and ϕ_o.

If $\theta_o = \phi_o = 0$ the beam is broadside to the ring and the pattern is simply

$$g_1(\theta,\phi) = J_o\left(\frac{2\pi a}{\lambda}\sin\theta\right) \qquad (9)$$

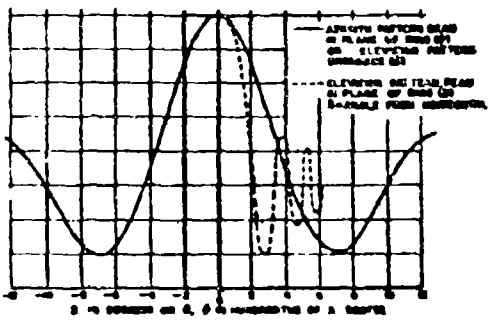


FIGURE 3 FAR FIELD VOLTAGE PATTERNS FOR SINGLE RING MODE 1 ANTENNA

If $\theta_o = \frac{\pi}{2}$, $\beta = 0$ the beam is in the plane of the ring and the pattern is

$$g_1(\theta,\phi) = J_o\left\{\frac{2\pi a}{\lambda}\left[(\sin\theta\cos\phi-1)^2 + (\sin\theta\sin\phi)^2\right]^{\frac{1}{2}}\right\} \qquad (10)$$

If we further assume $\phi = 0$ and determine the elevation pattern we have

$$g_1(\theta,0) = J_o\left[\frac{2\pi a}{\lambda}(\sin\theta-1)\right] = J_o\left[\frac{4\pi a}{\lambda}(\sin^2\frac{\theta}{2})\right] \qquad (11)$$

While $\theta_o = \frac{\pi}{2}$, $\phi = 0$, if we let $\theta = \frac{\pi}{2}$ we obtain the azimuth pattern.

$$g_1(\tfrac{\pi}{2},\phi) = J_o\left\{\frac{2\pi a}{\lambda}\left[(\cos\phi-1)^2 + (\sin\phi)^2\right]^{\frac{1}{2}}\right\}$$

$$= J_o\left(\frac{4\pi a}{\lambda}\sin\frac{\phi}{2}\right) \qquad (12)$$

Finally if we consider $\phi_o = 0$, $\theta_o = 0$ the beam is directed to a general elevation angle. If we let $\phi = 0$, an elevation cut is obtained.

$$g_1(\theta,0) = J_o\left[\left(\frac{2\pi a}{\lambda}\right)(\sin\theta_o - \sin\theta)\right] \qquad (13)$$

These patterns are shown in Figures 4, 5.

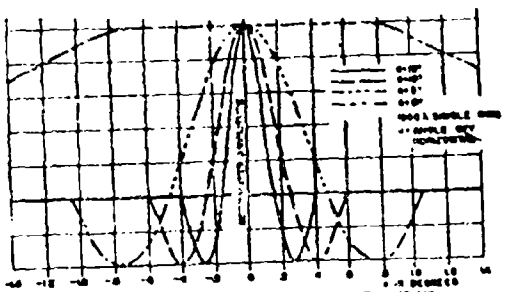


FIGURE 4. FAR FIELD VOLTAGE PATTERNS (ELEVATION) FOR SEVERAL BEAM DIRECTIONS

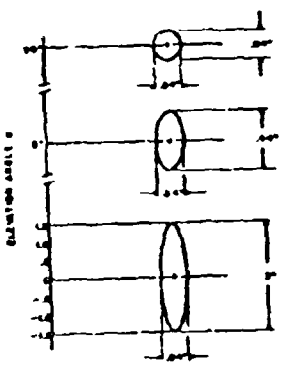


FIGURE 5. BEAM SHAPE AS A FUNCTION OF ELEVATION ANGLE FOR A 1000 λ SINGLE RING ANTENNA

The effect of beam broadening as a function of elevation angle of the beam is evident.

For multiple rings the patterns of several single rings are added as in Equation (7). The basic building block for a multiple ring array is a zero order Bessel function - a function that will not produce echelon lobes when large ring spacings are considered.

In pattern synthesis with ring arrays there are at least three variables that can be adjusted. These are relative amplitude of illumination of the various rings, relative ring radii, and beam direction of the individual rings.

When the number of rings used is small, there are minimum side lobe levels that can be achieved. It has already been noted that with a single ring there is a side lobe of 8db. For a two-ring array, side lobes of 12.3db have been achieved. For a three-ring array the maximum side lobe was reduced to 15.5db. Figures 6 and 7 show a two- and a three ring pattern respectively. As the number of rings increase the side lobe limitation disappears, and it becomes possible to create patterns with arbitrarily low side lobes. However selection of ring radius and amplitude of illumination for the various rings becomes a problem. In one case of array synthesis where the elements were to be directive, a Gaussian quadratures formula was used to determine the radii. Such a formula gives an excellent approximation to the desired function for a restricted angle. Outside this angle the element directivity is utilized. The use of such a formula alleviates the problem of selection of ring radii and amplitudes of illumination.

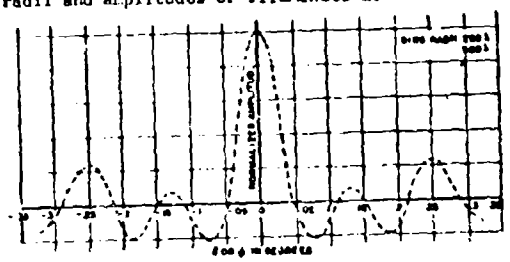


FIGURE 6. FAR FIELD VOLTAGE PATTERN (ELEVATION) FOR A TWO RING ANTENNA

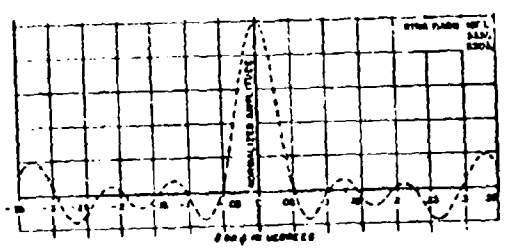


FIGURE 7. FAR FIELD VOLTAGE PATTERN (ELEVATION) FOR A THREE RING ANTENNA

Pattern Computation:

The patterns shown were generated with the aid of an analog computer, on which Bessel's equation of zero order may be easily solved. The computer circuit used is shown below, with the equation it solves.

$$\frac{d^2R}{dt^2} + \frac{1}{t}\frac{dR}{dt} + k^2R = 0$$

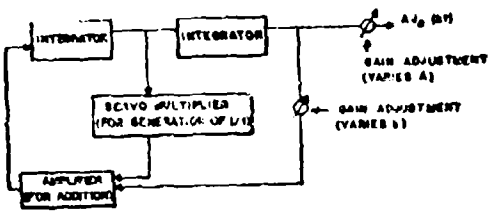


FIGURE 8

For a multiple-ring pattern one circuit is required for each ring. Varying one potentiometer has the effect of changing the strength of illumination, and varying the other has the effect of changing the ring radius. Effective beam direction is changed by adjusting initial conditions on the integrators.

It is interesting to note that this circuit also generates in function $J_1(u)$ and $\frac{J_1(u)}{u}$.

Number of Elements per Ring

An analysis by Knudsen[4] shown that for a single ring the pattern equations can be written in different forms, depending on whether the number of elements in the ring is even or odd.

N even

$$G = J_0\left(\frac{2\pi\rho}{\lambda}\right) + 2\sum_{q=1}^{\infty} J_{qM}\left(\frac{2\pi\rho}{\lambda}\right)\cos\left(\frac{\pi}{2} - \phi\right)2qM$$

N odd

$$G = J_0\left(\frac{2\pi\rho}{\lambda}\right) + 2\sum_{q=1}^{\infty} J_{qM}\left(\frac{2\pi\rho}{\lambda}\right)\cos\left(\frac{\pi}{2} - \phi\right)2qM$$

$$- 2\sum_{q=0}^{\infty} J_{(2q+1)M}\left(\frac{2\pi\rho}{\lambda}\right)\sin\left(\frac{\pi}{2} - \phi\right)(2q+1)M$$

where M = number of elements per ring.

In both expressions the leading term is the far field pattern factor for a single ring antenna with an infinitely large number of elements. The remaining terms are correction terms whose magnitude depend on the number of elements per ring, the ring radius, and the beam position. If the number of elements is large, only the lowest order correction term will be important. Even this term will be small if the order of the Bessel function is large in comparison to the argument of the function; that is $M \geq \frac{2\pi\rho}{\lambda}$. The maximum value of ρ is 2a or twice the radius of the ring. (This occurs for $\theta = \theta_0 = \frac{\pi}{2}$ and $(\phi - \phi_0) = \pi$.) When $\rho = 2a$, M must be larger than $\frac{4\pi a}{\lambda}$ and the spacing of the elements around the ring should be of the order of $\frac{\lambda}{2}$. This spacing agrees with the $\frac{\lambda}{2}$ criteria used in linear arrays to insure the existance of a single, well-defined maximum in the far field pattern.

If the beam is confined to higher elevation angles, the spacing can be increased slightly, since the value of ρ decreases. This fact is also pointed out by Ta-Shing Chu[11] in an article concerned with the use of circular arrays to obtain omnidirectional patterns.

Summary

In an environment requiring hardened antennas and functional flexibility, the ring array offers considerable promise. The array is flat and should therefore be easier to harden than forms with large vertical extent. Being an array, it has all the flexibility of other arrays, and in addition it has an azimuthal symmetry that is particularly useful in cases where it is necessary to scan or track 360° in azimuth. A narrow azimuth, shaped elevation beam for scanning is easily formed by proper phasing of the elements. Finally, a ring array with rings of greatly different radii has no wide angle diffraction lobes such as those that arise in building uniform rectangular arrays with widely spaced elements. Thus, for large arrays the number of elements can be systematically reduced by placing them in widely spaced rings.

References

(1) H. Chireix, "Antennas a resonnement zenithal reglint."
Onde Elect. vol 15, pp 440-456; 1936.

(2) W. W. Hansen and J. R. Woodyard, "A new principle in directional antenna design."
Proc. IRE vol 26, pp 333-345; March 1938.

(3) W. W. Hansen and L. K. Hollingsworth, "Design of 'flat-shooting' antenna arrays."
Proc. IRE, vol 27, pp 137-143; Feb. 1959.

(4) H. L. Knudsen, "Radiation from ring quasi-arrays."
IRE Transactions PGAP, pp 452-472; July 1956.

(5) Andrew Alford, Consulting Engineers, "High frequency steerable antenna."
pp 8-32 MC ASTIA No. AD 33342.

(6) Howard Q. Tottan, Antenna Beam Scanning
ASTIA No. AD 306302.

(7) H. Poritsky, Field Calculations of Radar Systems With Electrical Scanning
G.E. TIS No. 59GL10.

(8) H. Poritsky, Far Fields of Ring Distribution
G.E. TIS No. 60 GL 32

(9) A. E. Blume, Radiation Pattern of Circular Arrays
G.E. TIS No. M57APS121.

(10) Allan C. Schell and Edmund L. Bouche "A concentric loop array"
1958 IRE Wescon Convention Record.

(11) Ta-Shing Chu, "On the use of uniform circular arrays to obtain omnidirectional patterns."
IRE Transaction on Antennas and Propagation, October 1959.

TRACKING TRAINING RESEARCH AT THE U. S. NAVAL TRAINING DEVICE CENTER

By: Mr. Gene Micheli, Human Engineering Department, U.S. Naval Training Device Center

Skills called for in military situations have characteristically tended to be of a perceptual-motor nature; for example, tracking a radar blip, steering a submarine, firing a rifle or driving a tank. Common to a wide variety of these skills is the task of tracking, that is, a continuous perceptual-motor coordination, or in other words, following and "lining up" a system with some other system.

Tracking tasks may be grouped according to the functions required of the operator, as follows:

I. Plotting or following tasks
II. Flexible gunnery
III. Vectoring
    A. Radar traffic control
    B. Ground and air controller intercepts
    C. Aerial fixed gunnery
IV. Vehicular control
    A. Fixed wing aircraft
    B. Rotary wing aircraft
    C. Submarine depth control
    D. Submarine and ship steering
    E. Ground vehicles
    F. Remote controlled vehicles

The categories, while not completely exclusive from one another, show the various types of manual tracking.

This paper is a review of a continuing program of human engineering and training research in human tracking behavior which is conducted by the U. S. Naval Training Device Center. The purpose of the research program which is reviewed here is to improve human military performance through devising techniques and devices for training students in the development of tracking skills. Participating in this research program, which is planned, organized and supervised by the U. S. Naval Training Device Center, are various academic and industrial organizations. This paper traces the evolution of this research as a function of both changes in experimental techniques and operational requirements.

The earliest studies in this research program employed the traditional scientific technique of holding most of the variables in the tracking situation constant while systematically varying the particular factor under investigation in order to determine its effect on tracking performance. This was the technique used by State University of Iowa researchers from 1945 through 1947 to improve flexible gunnery performance through the design and use of specific training equipments.

These studies were followed by extensive investigations conducted by The Johns Hopkins University and University of Rochester psychologists on the effects of the location and design of displays and controls on human tracking performance. Studies were also done on the effects that various radar control settings, target characteristics and target loads would have on the performance of radar tracking tasks.

Information of the following nature resulted from these studies (Johns Hopkins and Rochester):

a. Controls should be located so that when the motion of the hand in moving the control is in one direction, the display moves in the same direction.

b. For a range of target sizes from 1/8" to 7/8", practice with a target of one size has the same training effect as practice with larger or smaller targets. With this range of targets, a skill learned with one size of target transfers completely to a new task using a different size of target.

c. The optimal gear-ratio between a crank and the cursor of a typical radar PPI depends mostly upon the angular distance through which the cursor moves to get on the target.

Today, in retrospect, we can say that these earlier investigations studied relatively elementary and basic skill. Yet, they met the needs of their time and provided valuable results and applications for the development of training devices and techniques.

As military technology advanced, the principle tracking requirement became one of vehicular control. In this situation the human element performs in a continuous closed-loop system. Research was begun, and is continuing, related to the dynamic features of this closed-loop situation. In these researches many variables are permitted to operate simultaneously, in contrast to the earlier studies wherein only one variable was manipulated at a time. An important result of a statistical technique which

makes this approach possible, the analysis of variance, is the determination of the interaction effects among the variables; that is, the relationship among them as they all vary at the same time.

A study was made by Electric Boat Company on the extent and character of the simulated dynamics necessary to develop skilled submarine operator performance. The purpose was to determine how closely a simulator must correspond to operational equipment, in this case a one-man control of a high speed submarine in course and depth. The variations in training effectiveness obtained under five different levels of fidelity of simulation were demonstrated in a controlled situation for three major tasks: depth changing only, course changing only, and simultaneous depth and course changing. One group of men was trained at each level of simulator complexity. Following training, all five groups were tested for proficiency in controlling the most complex simulator, i.e., the one which most closely resembled an actual submarine. (An actual submarine was not used, the assumption being made that the simulator was sufficiently like the real thing that proficiency in controlling the simulator would be a direct indication of proficiency in controlling a submarine.)

From this experimental design it was possible to determine how much benefit was derived from training on each of the four simplified simulators. The results showed that training on a relatively inexpensive and simple submarine simulator permits trainees to switch (or transfer) effectively to a much more complex and expensive simulator. There were no statistically significant differences between the groups trained on various degrees of simulation. However, there was a general trend for proficiency to be greater for the groups trained on the more complex simulators.

For training device design, this study provides valuable information on the degree to which an operational situation should be simulated for effective training, considering in addition, the factor of cost. The data suggest that while training effectiveness increases with increasing cost, it does so in proportionally smaller amounts. Therefore, a few dollars added to the cost of simple simulators tend to increase training effectiveness to a much greater extent than the same amount added to the cost of a more complex and expensive simulator.

A series of experiments was conducted by Ohio State University to investigate the role of the human operator as a sensor in a complex man-machine system with the objective of determining design requirements for tracking training devices. Two of these experiments were concerned with the effects of augmented feedback on the acquisition of skill in a tracking task. (Augmented feedback is defined as redundant feedback presented during training in addition to the cues provided by the tracking task itself. It is used even though the feedback from the task itself is adequate for high-level performance.)

In the first study, the investigators sought to determine the most efficient range from the target at which to introduce the augmented feedback. Four conditions were used: (a) a broad error tolerance group that received the augmented feedback (a click amplified through a loudspeaker) even when the tracking cursor was considerably off the target blip (the permissible error was broad enough to allow augmented feedback to be received about 68% of the time at first); (b) a medium error tolerance group that received augmented feedback approximately 51% of the time at first; (c) a narrow error tolerance group that received augmented feedback about 34% of the time at first; and (d) a group that received no augmented feedback at all.

It was found that the group which did not receive any augmented feedback at all gradually improved a little with practice. (After all, it did have some feedback -- the trainees could see how well they were tracking.) This group did not, at any time, attain the tracking proficiency reached by the three groups that did receive augmented feedback. The group with a broad error tolerance performed best at first, but its improvement soon tapered off. Tracking performance of the medium error tolerance group was not as good, at first, as that of the broad band group, but it eventually surpassed it. Similarly, the initial tracking performance of the narrow error tolerance group was the poorest of the three groups receiving augmented feedback, but its improvement continued the longest, until its tracking proficiency was the best of all the groups.

The implications of these findings are: (1) augmented feedback increases level of tracking performance; and (2) the degree of tracking error permitted within which augmented feedback can be received influences the rate at which performance improves.

In the second study, tracking performance was determined as a function of both the error tolerance band around the target and the schedule of augmented feedback, that is, a fixed amount (not responsive to increase in skill)

versus an amount that increased when skill in tracking increased. An interaction occurred between the size of the error tolerance band around the target and the schedule of augmented feedback. This interaction resulted in higher levels of tracking skill for: (a) trainees who had a narrow error tolerance band in which to receive augmented feedback but with a schedule that permitted the amount of augmented feedback to increase with increase in tracking skill and; (b) trainees who had a broad error tolerance band within which to receive augmented feedback but with a schedule that did not permit the trainee to receive more than a specified amount regardless of skill level. Thus, in selecting the amount of augmented feedback for a particular tracking training situation, it is important that both size of error tolerance band and schedule of augmented feedback be considered, as they do not influence the acquisition of tracking skill independently.

A preliminary investigation of tracking training by Dunlap and Associates resulted in a categorization of tracking tasks, the development of a descriptive model of tracking behavior designed to encompass all tracking tasks and the development of a group of hypotheses concerned with the transfer of training in vehicular tracking tasks.

As a follow-up to this area, by applying the preliminary findings to a representative tracking task, namely, submarine depth control, the researches found that several hypotheses previously formulated about tracking training were verified, while others were rejected. The major findings indicate:

a. The most important single factor in acquiring tracking skill appears to be practice. Thus, high levels of tracking skill may be developed relatively independently of intellectual comprehension of the dynamics of the tracking system.

b. Wide variations of the system equations in the equipment did not affect relative tracking skill to any significant degree. That is, large changes in the tracking task did not appreciably affect tracking performance.

These findings permit the generalization of recommendations to any vehicular tracking task, including submarine depth-keeping and steering. The recommendations are: (1) trainees should be given as much practice on a training device as possible in preference to lectures or classroom work; (2) training by lectures is of limited value and should be used only to provide an adequate insight into the over-all purpose and character of the tracking task; and (3) the training device does not have to simulate the complete system in order to be effective. Adequate simulation of the system dynamics should be done but precise simulation of controls, displays and task environment is not necessary.

It should be noted that, since tracking performance was only slightly affected by sizeable changes in the system equations, significant positive transfer may occur from one complex tracking task to others.

In other words, the researchers began to suspect that tracking skill was not really specific to particular systems, but included both general and specific components. Increases in skill with training appear to depend not only on mastering a given vehicle, with its particular indicators, controls, and dynamic characteristics, but also on the growth of a general psychomotor skill, resulting in better operation of a wide range of vehicles. It was believed that such a general tracking skill might best be developed by an inexpensive trainer designed to provide a wide range of tracking systems, that is, a general tracking trainer.

Experiments were conducted to determine the existence and nature of a general tracking skill or skills. One problem studied was the assessment of the degree of transfer from one simple general program of tracking training to three different specific tracking systems. The equipment gave practice on slow and fast one-dimensional tracking systems, for which stability was varied, becoming more unstable as skill increased. Then the trainees were tested on three conditions: (1) stable submarine-like condition; (2) unstable submarine-like condition; (3) aircraft condition (jet aircraft altitude control system with disturbance functions). (These systems were simulated on an analogue computer.)

The results demonstrated that the general tracking trainer established a considerable amount of skill which was transferred to the three test tasks, confirming the hypothesis that general tracking training produced important increases in skill in the control of systems having widely different dynamic characteristics. These increases in skill are probably equal to or better than that which would occur with specific training on the test (operational) systems. It is surmised that general training gives an additional profit in that a skill developed through general training has a very broad base, and might in the long run prove more valuable to, say, a pilot who will fly many varieties of aircraft in his lifetime than would any specific tracking skill, no matter how highly developed it might be.

Evidence confirms that vehicular training has very wide transfer value. Whether this is truly due to a cluster of human abilities that could be called a generalized tracking ability requires further validation. However, if further work does validate this hypothesis, it would permit the development of tracking trainers of a more universal nature with consequent wide utility in teaching fundamental tracking skills. Conversely, the expensive development of specific tracking trainers could be reduced.

We saw how in the past the research conducted has changed due to the introduction of new experimental techniques (the use of the statistical technique known as analysis of variance and the use of analogue computers) and how operational requirements have resulted in gunnery tracking giving way to vehicular control as the principal subject of investigation.

In the future, much effort will be exerted toward eliminating the human as a precise and continuous tracker. Machines will be designed which can track more precisely and respond more quickly than any human operator. Such machines will take over many of the functions now being performed manually; for example, for submarines, automatic depth-keeping and depth-changing at desired pitch rates; for ships and submarines, automatic steering; for air combat and anti-air warfare, automatic tracking of targets, computing of intercept courses and firing of missiles. While the advent of automation in military science makes the roles of the human being in this closed-loop situation less of a controller, the human's role as a sensor becomes more important. This is due to the fact that no mechanical or electrical device in the foreseeable future will match the eye's sensitivity, resolution ability, or ability to organize and interpret a wide variety of inputs (for example, color, brightness, texture, form, etc.). Therefore, when the system inputs are highly complex and are critical to the success of the tracking task, the human will supply the inputs to the system. For example, the control of ground vehicles and the tracking of ground targets, either visually, with television or with radar, are tasks which will probably continue to involve human tracking.

Consequently, future research plans will concern themselves with the sensory role of the operator. In the development of this skill, attempts will be made to form a basis for programming training equipment to vary in difficulty as a function of the increasing skill of the trainee. Work will be done on the programming of tracking trainers to give them the capability of automatically making the task more difficult for the operator as he acquires a greater skill on the task. One way this will be attempted is by having the size of a circle which circumscribes the reticle of a CRT be a function of the operator's skill. As the follower (cursor) approaches the reticle, the training device will automatically decrease the size of the circle.

Information obtained by research with this technique has implications for the use of motor-skill teaching machines, which like their verbal teaching machine counterparts, would have the capability of progressively training an operator. These teaching machines must be able to stop and repeat material that has not been well learned. In addition, they have to take the individual differences among trainees into account and also be able to provide feedback to the trainee to take the place of an instructor by defining acceptable performance levels.

Another area which will be studied in the future is the usefulness of a "contact analogue" device for training in the control of land vehicles, surface ships and aircraft. This electronic analogue will present an abstract representation of the "real world" on a screen so as to provide a non-programmed visual display to which a trainee would respond as if to real situations.

The purpose of the research program at the U. S. Naval Training Device Center is to improve the human military tracking performance that is required by current operational situations by increasing the effectiveness of training devices and techniques. Its objective can, of course, be met only to the extent to which the research findings are incorporated into the design of tracking training devices.

It was shown how the character of this tracking research program reflects the changes in experimental techniques, the availability of advanced research equipment and the ever-changing nature of the task of the military man as military science becomes more advanced.

ON THE TRACKING POTENTIALITIES OF A DOPPLER RATE MEASURING SYSTEM

BY: Duane C. Brown, RCA Missile Test Project, Patrick AFB

CONVENTIONAL USE OF DOPPLER

The first tracking system in this country to exploit the doppler effect was the DOVAP (DOppler Velocity and Position) system developed in the mid 40's by BRL. Descendents of DOVAP (UDOP, EXTRADOP, HYDERDOP) are in use at the Atlantic Missile Range and elsewhere. A common property of such systems is that they cumulate doppler cycles and hence may be considered to be integrating systems. Positions are determined by the intersection of the quadric surfaces generated when cycle counts are converted to ranges, range sums or range differences. One transmitting and at least three receiving stations are required for a minimal solution; more than three receiving stations will permit a least squares solution.

THE STARTING POINT PROBLEM

In the conventional reduction, it is necessary that an externally determined point be provided for zeroing the cycle counts. This dependency upon other systems for a starting point has seriously compromised doppler systems, especially since fresh starting points are required after each signal dropout of significant duration. Further difficulties arise from the fact that errors in starting points tend to propagate unfavorably and may be magnified one-hundredfold towards the end of coverage. Accordingly, considerable effort has been expended to devise means to circumvent the need for externally determined starting points.

INTERNAL DETERMINATION OF STARTING POINTS

Notable solutions to the starting point problem were developed by Guard (1) and Willman (2) who exploited the observational redundancy which exists when more than three receiving stations acquire data. Instead of determining each trajectory point independently, the solutions determine all trajectory points simultaneously in a massive least squares solution which carries the coordinates of the starting point as additional unknowns. For a total of n trajectory points, a system of normal equations of order $3n+3$ results. Though of seemingly formidable dimensions, the normal equations turn out to be so patterned with zero elements that their inversion is readily accomplished by partitioning; the over-all computations increase only linearly with n. Willman's solution was employed with considerable success at AMR for the reduction of Redstone trajectories. On the other hand, the results for the longer range Jupiter trajectories were unsatisfactory, for the solution tends to be indeterminate as the portion of the trajectory utilized approaches linearity. The Redstone, which is normally tracked from shortly after launch to just before impact, provided sufficient variation in geometry to produce a satisfactory solution, whereas the approximate linearity of the observed portion of the Jupiter trajectory produced very poor results. Because of this weakness, the solution has proven to be of limited value at AMR.

SIX STATION RATE-ONLY SOLUTION

An entirely different approach to the starting point problem was proposed by Bruhns (3) who exploited the time derivative of the range-sum equation. In this solution, doppler cycles counted over a short interval are converted to doppler frequencies and thence to range-sum rates. Rates from six stations provide six equations involving the coordinates of the transponder X, Y, Z and its components of velocity $\dot{X}, \dot{Y}, \dot{Z}$ as unknowns. Because of the differentiation of the range-sum equations, the coordinates of the starting point disappear and hence are not required. An error analysis of this solution proved to be quite disappointing. In fact, the solution for many of the points considered appeared to border on indeterminacy. For this reason, it was judged to be unfeasible and was never employed at AMR in actual reductions.

REFERENCE FREQUENCY TRANSMISSION

One of the chief practical limitations of DOVAP and other existing CW systems has been their dependency on the transmission to the receiving stations of the reference frequency of the transmitter. The DOVAP reference frequency at AMR, for instance, is transmitted by cable and by microwave to stations as far as 150 miles from the transmitter. The expense of such transmission for several stations well-distributed over a region many hundreds of miles across would undoubtedly be prohibitive. One obvious way around this problem would involve having each station both transmit and receive. However, this would entail the use of a separate transponder for each station (or else an elaborate transponder sharing technique); thus to avoid an undue weight penalty the number of stations capable of simultaneous operation would necessarily be quite limited. Another possible solution might eventually result from the use of VLF transmission, provided that stabilities an order of magnitude better than the 1 in 10^{10} reported in (4) can be attained. What is perhaps the cleanest solution of all is also presently beyond the state of the art: the use of an

individual and sufficiently stable frequency standard at each receiving station. The long and short term stabilities needed for modern tracking requirements are both about one part in 10^{11}; this is tenfold better than the stabilities of the best standards commercially available. On the other hand, such effort is being expended throughout the country on the development of improved frequency standards that the attainment of sufficient stability appears to be only a matter of time.

RECONSIDERATION OF RATE-ONLY SOLUTION

In late 1959, well over two years after the initial study and discarding of the rate-only solution discussed earlier, the writer was induced by potential progress in the development of ultra-stable frequency standards to reconsider the rate-only approach to the utilization of doppler. An initial stimulus was provided by the proposed NASA experiment to check the general theory of relativity by means of a satellite-borne atomic clock having a frequency stability purportedly approaching a few parts in 10^{11} and weighing less than 50 pounds. Also, it had been suspected (incorrectly) that the rate-only solution would approach indeterminacy as all receiving stations were to approach the surface of a plane; therefore, it was reasoned that the rate-only solution might make a better showing if the stations were more widely distributed on the surface of the earth. Final impetus was provided by the consideration that the use of independent frequency standards would permit considerable flexibility in locating stations and would allow the use of an unlimited number of stations.

RATE-ONLY SOLUTION FOR SPHERICAL SYSTEM

The simplest form of a rate-only system arises when transmission is from a frequency standard aboard the vehicle being tracked, for then straight range rates are obtained at each receiving station. Such a system may be called a spherical range-rate system. Elliptical or hyperbolic systems would result if a transponder were to replace the frequency standard in the vehicle so that range-sum rates or range-difference rates were measured at each station. If a receiving station were located in the immediate vicinity of the transmitter, the accuracy of an elliptical or hyperbolic system would be very nearly the same as that of the corresponding spherical system. Accordingly, it largely suffices to consider accuracies attainable from a spherical system. The observational equation for a spherical system is obtained from the time derivative of the ranging equation:

$$r_i = [(X - X_i)^2 + (Y - Y_i)^2 + (Z - Z_i)^2]^{\frac{1}{2}} \quad [1]$$

wherein r_i is the range from the vehicle at X, Y, Z to the i^{th} tracking station at X_i, Y_i, Z_i. Differentiation of r_i gives

$$\dot{r}_i = \frac{X - X_i}{r_i}\dot{X} + \frac{Y - Y_i}{r_i}\dot{Y} + \frac{Z - Z_i}{r_i}\dot{Z} \quad [2]$$

which expresses the measured range rate $\dot{r}_i$ as a function of the vehicle's components of position and velocity. Six stations measuring $\dot{r}$ thus provide six equations for the determination of $X, Y, Z, \dot{X}, \dot{Y}, \dot{Z}$. When more than six stations provide data, an overdetermined solution is possible.

RESULTS OF NUMERICAL STUDY

In order to assess the accuracies to be expected from a spherical range-rate system, a hypothetical system was postulated to exist on the Atlantic Missile Range with receiving stations at the locations indicated in Figure 1.

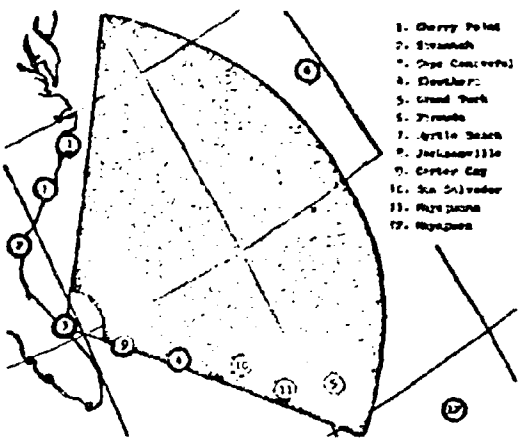


Figure 1. Geometrical configuration encountered in numerical study.

Selected points on the trajectory which was assumed to be observed by the system are listed in Table 1 for the case of a 90° launch azimuth from Cape Canaveral. The Z axis is normal to the reference spheroid at Cape Canaveral and is positive upwards. The Y axis is tangent to the meridian and is positive to the north; the X axis is tangent to the prime vertical and is positive to the east. The trajectory was assumed to be launched at azimuths ranging from 30° to 135°, the general region of coverage being indicated by the shaded area in Fig. 1. The trajectory is a high performance one designed to provide a stringent test of any tracking system. Both six and twelve station solutions were considered. Stations 1 through 6

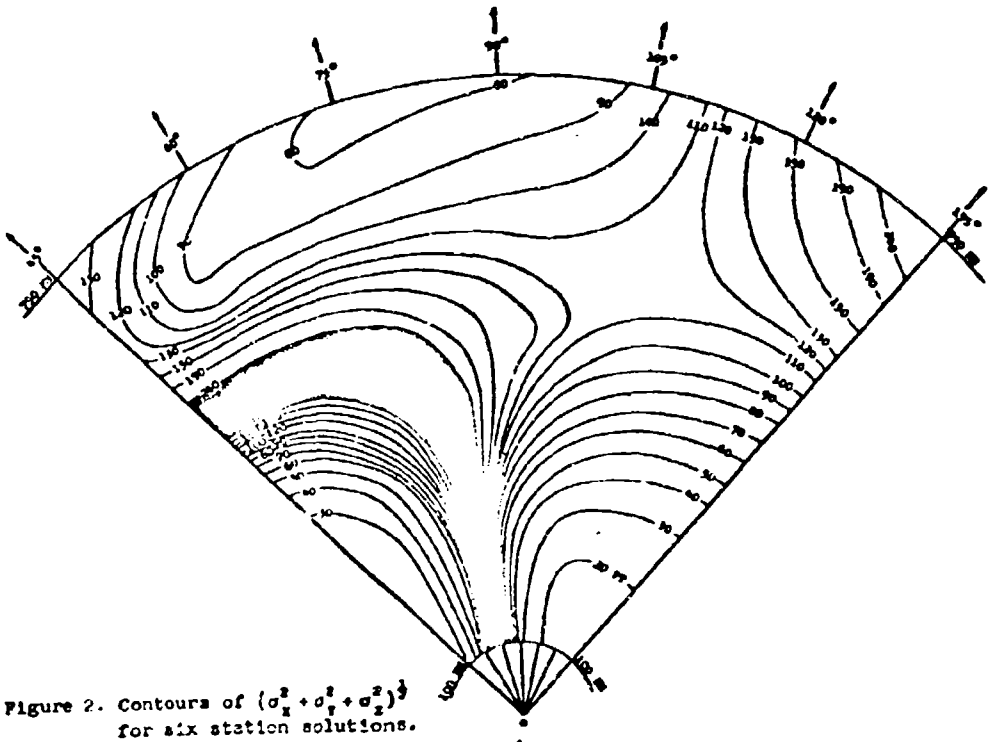


Figure 2. Contours of $(\sigma_x^2 + \sigma_y^2 + \sigma_z^2)^{\frac{1}{2}}$ for six station solutions.

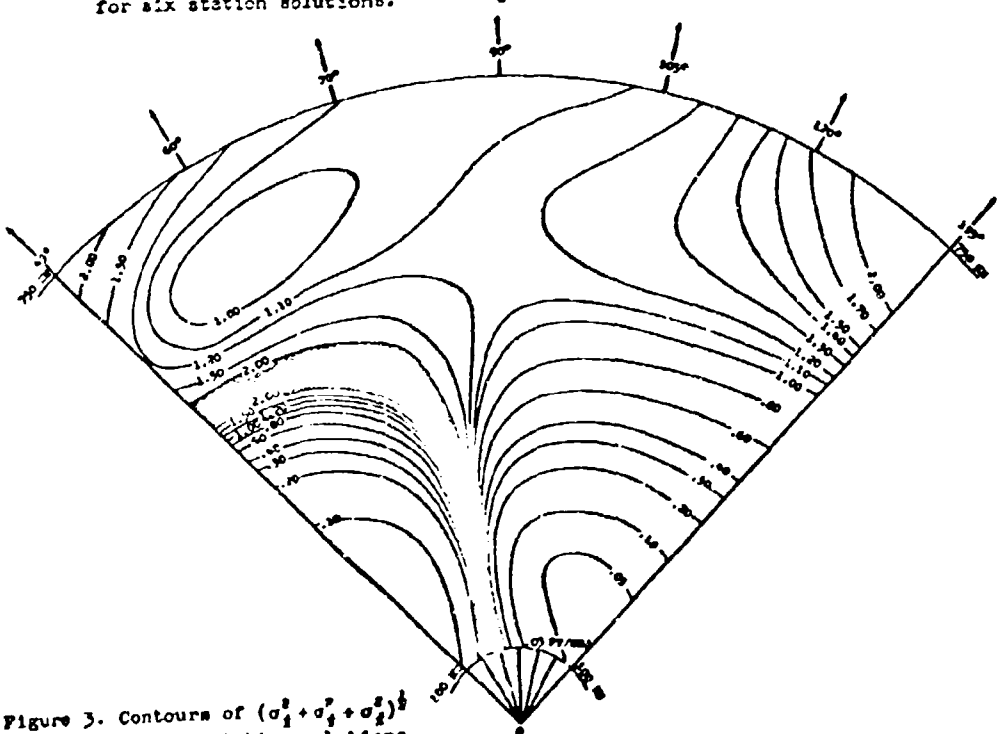


Figure 3. Contours of $(\sigma_{\dot{x}}^2 + \sigma_{\dot{y}}^2 + \sigma_{\dot{z}}^2)^{\frac{1}{2}}$ for six station solutions.

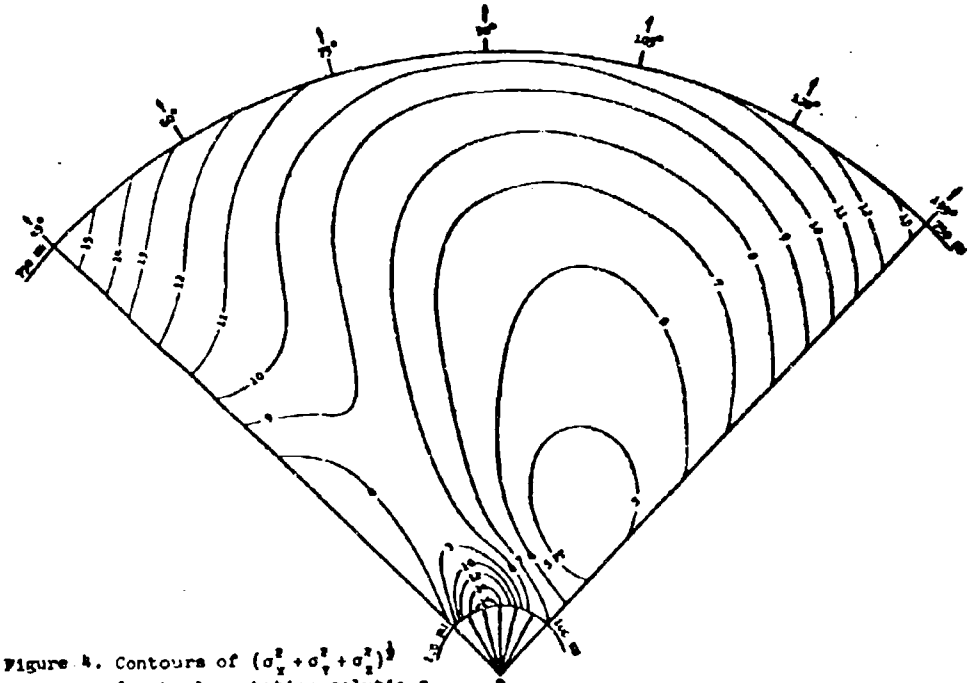


Figure 4. Contours of $(\sigma_x^2 + \sigma_y^2 + \sigma_z^2)^{\frac{1}{2}}$ for twelve station solutions.

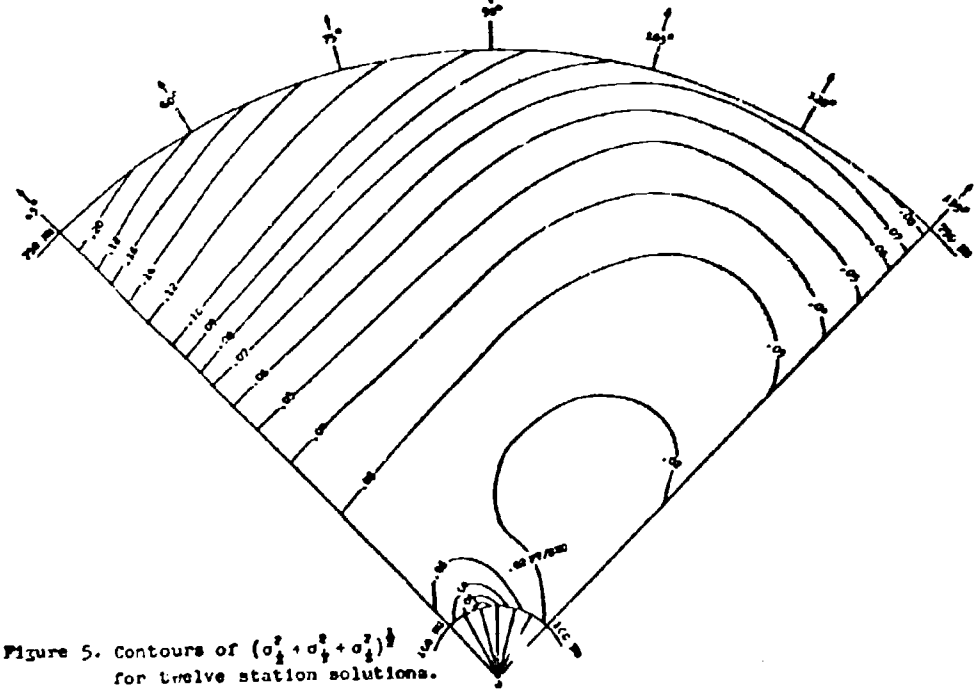


Figure 5. Contours of $(\sigma_{\dot{x}}^2 + \sigma_{\dot{y}}^2 + \sigma_{\dot{z}}^2)^{\frac{1}{2}}$ for twelve station solutions.

were employed in the six station solutions. The measuring accuracy (i.e., standard deviation) for range-rates was taken as 0.01 ft/sec for all stations. No error was attributed to the figure for the velocity of light in vacuo nor were survey errors considered. The standard deviations of position and velocity are summarized by the 'error' contours in Figs. 2-5. The shaded areas in Figs. 2 and 3 correspond to regions where the gradients of the error contours were too steep to permit delineation. They thus define regions of near indeterminacy. It is to be emphasized that no attempt was made in the numerical study to optimize the locations of the stations; the selection was based primarily on logistical considerations. Presumably, then, optimization would lead to significant improvements.

t sec	X ft	Y ft	Z ft	$\dot{X}$ ft/sec	$\dot{Y}$ ft/sec	$\dot{Z}$ ft/sec
200	602 566	0	559 982	10 532	0	6 007
220	824 377	0	680 574	11 675	0	6 067
240	1 070 565	0	809 240	12 943	0	6 200
260	1 343 182	0	929 166	14 368	0	6 422
280	1 646 593	0	1 060 670	15 995	0	6 761
300	1 984 900	0	1 200 722	17 809	0	7 255
320	2 354 520	0	1 352 426	20 155	0	7 963
340	2 794 691	0	1 521 315	22 978	0	8 993
360	3 269 700	0	1 715 685	26 722	0	10 554
380	3 875 451	0	1 950 352	32 297	0	13 173
400	4 615 594	0	2 262 204	43 490	0	19 036

Table 1. Hypothetical high performance trajectory used in numerical study.

IMPROVEMENT EFFECTED BY REDUNDANCY

The six station solution is seen from Figs. 2 and 3 to yield widely varying accuracies. Over a sizeable region the solution approaches indeterminacy; yet within 25 miles of this region, positional accuracies of better than 50 ft. and velocity accuracies of better than 0.5 ft/sec are attained. An investigator who by chance happened to select for study only points in the region of near indeterminacy would naturally conclude that a six station range-rate system is impractical. It now appears as if something similar to this occurred in the original investigation of Bruhn's modified DOVAP solution. In general, Figs. 3 and 4 would lead one to the conclusion that the postulated six station system would be competitive with existing tracking systems only for a limited volume of coverage. On the other hand, the improvement resulting from the twelve station system is seen from Figs. 4 and 5 to be most pronounced. The regions of near indeterminacy have been wiped out and replaced with well-behaved contours of low magnitude. Over most of the region of coverage, the positional accuracies are better than 10 ft. and the velocity accuracies are better than 0.1 ft/sec. This degree of improvement resulting from moderate redundancy was considered to be quite startling because it was contrary to all experience gained at AMR with other tracking systems. Almost any experienced data analyst is likely to conjecture that when the number of observational stations in an already well-balanced configuration is doubled, the improvement to be expected should not much exceed a factor of $\sqrt{2}$. Accordingly, appreciable effort was devoted to seeking an explanation for this apparently anomalous behavior of a range-rate system. The explanation ultimately arrived at hinged on the observation that the error contours for six station solutions are highly sensitive to the locations of the stations. In general, a six station configuration will produce fairly good accuracies only within a certain rather limited region; outside this region accuracies deteriorate rapidly. If the configuration is altered slightly, say by replacement of one of the stations, the error contours usually become markedly altered and the region of accurate coverage is shifted considerably. Hence, when many different six station configurations are possible, one is usually able to choose one which will produce good results for a given small area within the overall configuration. This, together with the fact that an overdetermined solution is necessarily superior to those resulting from any of the possible minimal combinations explains the rather remarkable improvement resulting from redundancy. It is felt that the major contribution of this paper is the disclosure that a nonintegrating doppler system measuring range rates only can provide extremely accurate trajectory data over a wide volume of coverage, provided that a moderate degree of redundancy is exploited.

REFRACTION CORRECTIONS

One advantage of a range-rate system is its relative insensitivity to refraction, the refraction for range rate being the time derivative of the refraction for range. The refraction correction $\Delta\dot{r}$ for a point outside an atmosphere having an index of refraction profile of the form $\mu = 1 + (\mu_o - 1)\, e^{-bh}$, μ_o being the index of refraction for altitude h=0, is provided by the following 'flat earth' formula

$$\Delta\dot{r} = \frac{\mu_o - 1}{b} (\csc E \cot E)\, \dot{E} \qquad (E > 10^\circ) \,. \quad [3]$$

in which E denotes the elevation angle and $\dot{E}$ its rate of change. Since $\dot{E}$ for stations observing at low elevation angles would ordinarily be quite small, say less than .005 radians/sec, and since the quantity $(\mu_o - 1)/b$ would rarely be in error by more than .05, the uncertainty in the correction $\Delta\dot{r}$ would

usually amount to only a few thousandths of a foot per second. Hence, potential system accuracies are not seriously compromised by refraction.

PARAMETERIZATION OF TRANSMISSION FREQUENCY

An obvious weak point of a spherical system is its dependence on the assumption that the frequency standard aboard the vehicle will maintain its static stability during and after a period of strong acceleration. A possible way around this difficulty is to regard the instantaneous transmission frequency as an unknown along with the instantaneous position and velocity. For this approach the basic doppler equation is put into the form

$$\dot{r}_i = c\left[\frac{\Delta f_i}{f_o} + \frac{f_o - f_i}{f_o}\right] + \Delta\dot{r}_{ii} + \Delta\dot{r}_2 + \Delta\dot{r}_3 \quad [4]$$

in which

- c = vacuo velocity of light,
- $\dot{r}_i$ = range rate at i^{th} station,
- f_i = reference frequency at i^{th} station,
- Δf_i = beat frequency at i^{th} station = $f_i - f_i'$ where f_i' is received frequency,
- f_o = transmission frequency,
- $\Delta\dot{r}_{ii}$ = refraction correction,
- $\Delta\dot{r}_2$ = special relativistic correction,
- $\Delta\dot{r}_3$ = general relativistic correction.

The relativistic corrections are discussed in a subsequent paragraph. In an ideal system, one would have $f_i = f_o$ and the beat frequency (which constitutes the actual measurement) would be equal to the doppler frequency. When f_o is unknown one may replace $\dot{r}_i$ in [4] by the above expression. The equations generated by seven stations are sufficient to determine f_o as well as X,Y,Z, $\dot{X},\dot{Y},\dot{Z}$. An adjustment may be performed when more than seven stations acquire data. A preliminary examination of this extended solution indicates it to be reasonably sensitive under conditions of moderate redundancy. For instance, the accuracies to be expected from the twelve station configuration considered earlier are for the 5^{th} point in Table 1:

σ_{f_o}	= .75 cps		
σ_X	= 3.53 ft	$\sigma_{\dot{X}}$	= .020 ft/sec
σ_Y	= 5.83 ft	$\sigma_{\dot{Y}}$	= .007 ft/sec
σ_Z	= 5.19 ft	$\sigma_{\dot{Z}}$	= .167 ft/sec

In this computation the frequencies f_i and f_o were taken as 10 kmc and the beat frequencies Δf_i were assumed to be measured to an accuracy of 0.1 cycles. The corresponding solution for the case in which f_o is assumed to be known perfectly gives

σ_X	= 1.89 ft	$\sigma_{\dot{X}}$	= .014 ft/sec
σ_Y	= 2.70 ft	$\sigma_{\dot{Y}}$	= .006 ft/sec
σ_Z	= 4.87 ft	$\sigma_{\dot{Z}}$	= .015 ft/sec

If further study should show that relative accuracies similar to the above are maintained over an extensive volume of coverage, the extended solution will permit pronounced relaxation of the stability of transmission frequency in many applications.

PARAMETERIZATION OF STATION FREQUENCIES

A more far reaching extension, which is also under study, is based on the premise that the frequency standards at the receiving stations may have inadequate long term stabilities (say 1 in 10^{10}) but satisfactory short term stabilities (say 1 in 10^{11}) over the period of tracking. The solution is based on the fact that the beat frequency at the i^{th} station for the j^{th} trajectory point is then functionally of the form

$$\Delta f_{ij} = g(f_i, f_o, X_j, Y_j, Z_j, \dot{X}_j, \dot{Y}_j, \dot{Z}_j) \quad [5]$$

In the ensemble of mn such equations arising from m stations and n trajectory points there will exist 6n unknown trajectory parameters (the X_j, Y_j, Z_j, $\dot{X}_j$, $\dot{Y}_j$, $\dot{Z}_j$) together with m unknown reference frequencies f_i and n unknown instantaneous transmission frequencies f_o. If an elliptical or hyperbolic system is employed, only a single unknown transmission frequency f_o need be considered, provided transponder stabilities are adequate. If m is greater than 7 and if n is sufficiently large, the total number of equations will exceed the total number of unknowns and a solution may be attempted. The normal equations for the minimum variance adjustment are of the order m+7n, but are amenable to solution, no matter how large n may be, because of the preponderance of zero elements in the coefficient matrix; as in Willman's solution for DOVAP starting point, the computations increase only linearly with n. If this approach should prove feasible, the primary requisite for the frequency standards will be reduced to one of high short term stability, short term here referring to the typical duration of tracking. The results of a feasibility study now underway will be published at a later date.

RELATIVISTIC CORRECTIONS

A range-rate system capable of observational accuracies of 0.01 ft/sec. will be affected quite significantly by both the special and the general relativistic effects on frequency. The special relativistic correction (time dilation) as applied to range rates is

$$\Delta\dot{r}_2 = -\frac{1}{2}\frac{v^2}{c} + \text{higher order terms} \quad [6]$$

and the general relativistic correction (caused by differences in gravitational potential) is

$$\Delta\dot{r}_3 = \frac{k}{c}\left(\frac{1}{\rho} - \frac{1}{\rho_e}\right) + \text{higher order terms} \quad [7]$$

In these equations

v = total velocity of vehicle,

k = gravitational constant = 1.4032×10^{16} ft^3/sec^2,

ρ_e = geocentric distance of receiving station,

ρ = geocentric distance of vehicle = $\rho_e + h$.

For velocities of 20,000 ft/sec. $\Delta\dot{r}_2$ = 0.20 ft/sec. which is 20 times the basic measuring accuracy assumed herein; for 30,000 ft/sec, $\Delta\dot{r}_2$ = 0.45 ft/sec. Values of $\Delta\dot{r}_3$ versus altitude (h) are listed in Table 2

h(nm)	$\Delta\dot{r}_3$ (ft/sec)
100	.019
250	.046
500	.096
1000	.152
2000	.248
4000	.362
8000	.472
∞	.674

Table 2. General relativistic corrections versus altitude.

Because of the significance of the relativistic corrections, it is entirely possible to reverse matters and design a test, quite different from that in the proposed NASA experiment, to check relativity. In its simplest form the test would rely on frequency stabilities of a few parts in 10^{11}. From ten to fifteen tracking stations would be required. Each trajectory point would be reduced independently with $\Delta\dot{r}_2 + \Delta\dot{r}_3$ being carried as an additional unknown in much the same manner as f_o was carried in the first extension of the range-rate solution (because the gravitational potential at all tracking stations would be essentially identical, the corrections $\Delta\dot{r}_3$ would, like the $\Delta\dot{r}_1$, depend only on the trajectory point). Under the assumption of frequency stabilities of 1 in 10^{11} the standard deviation of the determination of $\Delta\dot{r}_2 + \Delta\dot{r}_3$ for the point considered in the preceding paragraph would be 0.07 ft/sec. This independent, point-by-point solution would not permit the separation of the special and general relativistic effects; to isolate the general effect it would be necessary to stipulate the special effect. However, by comparing the values of $\Delta\dot{r}_2 + \Delta\dot{r}_3$ experimentally determined over the wide range of velocities and altitudes resulting from a vertical space probe with those computed from theory, one can evaluate the correctness of the combined theories. A more sophisticated solution along the lines of that in the preceding paragraph could be employed to counteract inadequate long term stabilities. Also, the incorporation of orbital constraints into the adjustment should lead to increased sensitivity. Perhaps the principal drawback of the present approach in comparison with that of NASA is the number of stations required. On the other hand, the stations would provide instantaneous determinations of the combined relativistic corrections as a function of altitude and velocity while simultaneously providing precise trajectory data. Furthermore, the observational residuals resulting from the adjustment would provide valuable checks on internal consistency and could be employed for data editing.

POSSIBLE GEODETIC APPLICATIONS

The sensitivity of a rate-only system to errors in survey increases with increasing velocity. For velocities of 25,000 ft/sec a survey accurate to about 1 part in 750,000 would be required to prevent significant dilution of the potential accuracies of a system measuring $\dot{r}$ to .01 ft/sec. In view of this, it would appear logical to reverse the problem by treating the coordinates of some of the stations as unknowns. The principal drawback which emerges is the unduly large number of stations required for an effective solution; from nine to twelve stations must be of precisely known location to begin with, and the unknown stations cannot be very far removed from the configuration of known stations. Although a rate measuring system, by itself, does not appear to be very promising as a geodetic tool, a preliminary study has shown that the combination of a rate system and a long focal length photogrammetric system, which photographs strobe flashes against the stellar background, can be quite effective. When rate measurements accurate to .01 ft/sec were assumed to be incorporated into a hypothetical six station photogrammetric net which had been designed to produce survey accuracies of about 1 part in 1,000,000, the combined system was found to produce almost doubled accuracies. Considerable further study is required, however, before the precise role that rate measurements are capable of assuming in geodesy can be adequately evaluated.

REFERENCES

(1) R. Guard, "A Technique for the Determination of DOVAP Start-Point," ORDBS-OM-T10, White Sands Proving Ground, Tech Memo. 428, July 1957.

(2) J. Willson, "A Least Squares Solution for DOVAP Position and Tie-in Point Considering Refraction," RCA Data Red. Tech. Report No. 42, Jan. 1958.

(3) R. Brune, "A Modified DOVAP Reduction," Army Missile Agency, Missile Firing Lab., July 1957.

(4) U.S. Naval Observatory, Time Service Notice No. 7, June 1959.

CALCULATION OF MEASUREMENT ERRORS BY A STATISTICAL TECHNIQUE

by
L. G. Lorson
Philco Corporation
Western Development Laboratories
Palo Alto, California

The accuracy of test systems and equipment is often calculated by using the summation of magnitudes method. Because most errors have a normal distribution, a statistical method of calculating accumulating errors is proposed; such a method will eliminate most of the weaknesses of the summation of magnitudes methods and will give a more accurate test system or equipment error. Further, the alternative statistical method shows that test system and equipment are more accurate than they appear under the summation of magnitudes method.

Requirements for precision equipment in our present missile and space programs have placed a heavy technological burden on the test system designer. The test system designer must design equipment and test systems capable of checking the performance characteristics of high precision equipment. To do this, the test system must be more accurate and precise than the equipment tested. The test system designer usually calculates the errors in the equipment to be tested, and designs a test system that is an order of magnitude more accurate.

The actual assignment of accumulating error, i.e., system accuracy, becomes the critical element of every test system design. It affects costs, delivery schedules, and even the feasibility of the program. The results of this calculation affects design, marketing, sales, and management in all technological areas. It critically affects military programs. The need for a meaningful calculation of accumulating errors is obvious.

Following are two different methods of computing system errors: First, summation of the positive and negative magnitudes of each contributing error. Second, a statistical calculation of accumulating errors. Both methods are based on identical test systems so that the results may be compared. Since the calculations are dissimilar, the individual errors are defined differently; however, their relation to each other is clear. The test system is shown in Fig. 1. The purpose of the test is to calculate the frequency response of a transmitter. The transmitter is modulated by a wide-band oscillator, the signal is detected in an ultralinear AM/FM test receiver, and the output of the modulating source and the test receiver are presented on a dual-trace oscilloscope.

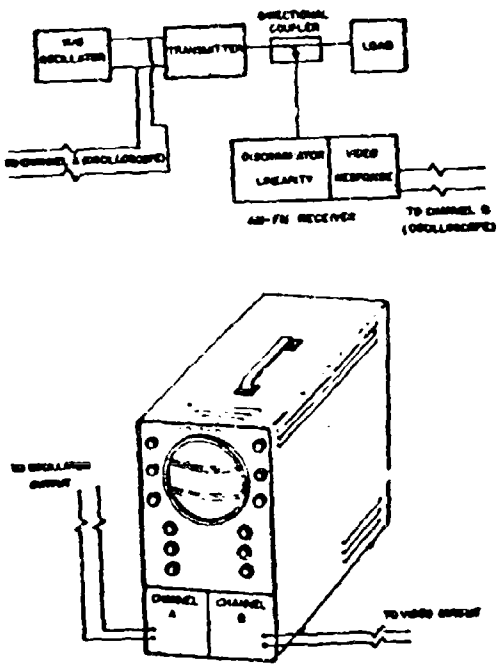


Figure 1 Test Set-up for Evaluating the Dynamic Operating Characteristics of the Transmitter (Shaded) Under Test.

For the test, the amplitude of the signal of the modulating source is set at a given level and the amplitude of the test receiver signal is simultaneously observed on the other channel. The frequency of the modulating source is changed and the new amplitude of the modulating signal is set to the previous level (Fig. 2). When this has been done, the amplitude at the output of the test receiver is noted and the differential amplitude is computed. From this, the frequency response of the transmitter is determined.

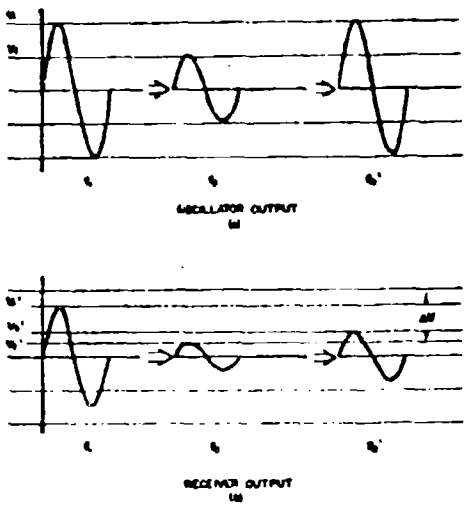


Figure 2 (a) Oscilloscope Pattern for Oscillator Output Shows Amplitude Reset to Original Amplitude. (b) Oscilloscope Patterns Illustrate Calculation of ΔV.

In the first calculation, each individual error is defined as a region within two definite points (Fig. 3). The error in re-setting the oscillator amplitude to its original level is defined as ± 2 per cent. Since the error involved in reading the differential amplitude (Δ A) involves reading two amplitudes, the error in reading Δ A is defined as ± 3 per cent. The author admits these limits are arbitrary; however, unless one determined these errors statistically, this would always remain a problem of engineering judgement. The error in the linearity of the discriminator in the AM/FM receiver is ± 1.5 per cent. The error in gain of the video section is ± 4 per cent. Because of the experimental hookup, the errors in the load and the modulating source do not contribute to the accumulating errors in the system. We shall assume the errors in the oscilloscope cancel. Even if this were not the case, these errors could be handled in the same manner as the other contributing errors.

The total system error (Fig. 3b) is given by

$$E_{T\pm} = \sum_{i=1}^{n} E_{i\pm}$$

where E_i = magnitude of each contributing error. In the example shown, the total system error is

$$E_{T\pm} = E_{1\pm} + E_{2\pm} + E_{3\pm} + E_{4\pm}$$

$$E_T = (\pm 1.5)+(\pm 2.0)+(\pm 4.0)+(\pm 3.0)$$

$$E_T = \pm 10.5\%$$

If the absolute magnitudes of the limits can be established, this method of determining the magnitude of accumulating error is perfectly valid, but the information it yields is limited. For example, there is no indication of the frequency at which an error of the value computed will occur. There also is no indication of which error ought to be changed, or which would be most economical to change. In addition, the calculation assumes the errors are fixed, i.e., that they can never be greater than the magnitude given. We know that transients or catastrophic failures can distort an individual reading. This is especially true when human error is introduced in the test setup.

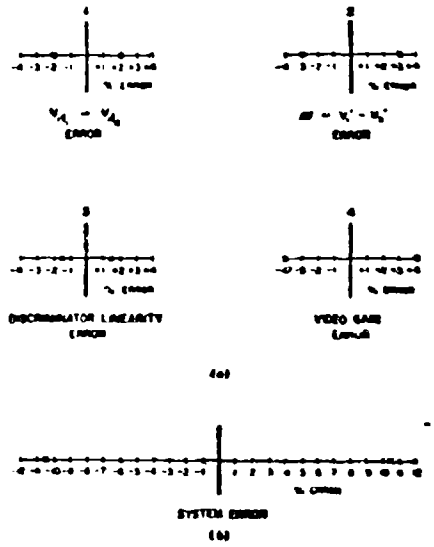


Figure 3 The Positive and Negative Magnitudes of Errors 1, 2, 3, and 4 are Summed to Give a Total System Error of ±10 1/2%.

This type of information can be likened to a doctor who tells his patient that there is a possibility he could die from a tonsillectomy, but neglects to tell him such a probability is 0.000001. One advantage of this type of calculation is its simplicity.

Basically, the method of calculating the test system error proposed by this paper will determine the distribution of the system error from the distributions of the individual errors. Since the statistical procedures and formulas are not new, no proofs of the formulas will be given. These proofs are found in any elementary book on statistics.

The first and most difficult step in calculating the system error is to determine the distribution of the individual errors. For the test system shown in Fig. 1 this is done as follows:

(1) For our example, the error in resetting the oscillator amplitude to its original level is ±2 per cent; such error has a normal (gaussian) distribution.

This does not mean that it is impossible to make an error greater than 2 per cent; most of the errors in resetting the oscillator are near zero, and only a small per cent of them exceed ±2 per cent. Let's assume that an error greater than ±2 per cent is made only one time in ten. In other words, 90 per cent of the time the error is between -2 per cent and +2 per cent. This is illustrated in Figs. 4 and 7a.

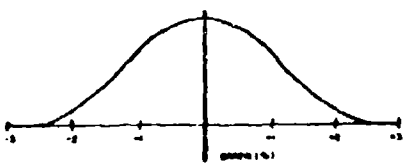


Figure 4 Normal Distribution of Error Involving Resetting Amplitude of Oscillator at f_2 to Amplitude Recorded at f_1. (Ninety Per cent of the Area Lies Between ±2%).

(2) The error involved in determining delta A or the differential amplitude was defined (summation method) to be ±3 per cent. In other words, the operator can determine the delta A to within ±3 per cent at least 90 per cent of the time. This error also has a normal distribution, which is shown in Fig. 7d.

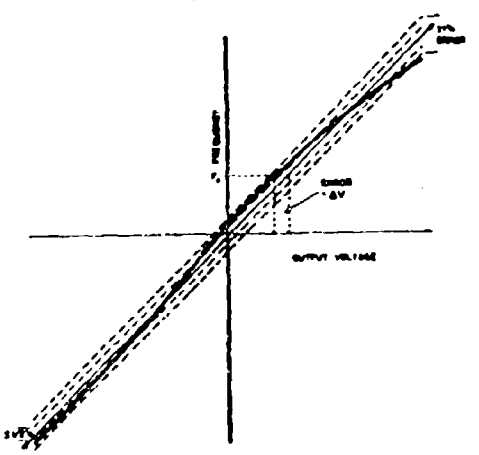


Figure 5 Linearity Curve - Plus and Minus 1% Error Lines Shown; Plus and Minus 1/2% Error Lines Shown.

(3) Since the error in the linearity of the discriminator is given by the deviation of a fixed plot relative to a straight line (Fig. 5), and since two parallel lines define the error, we can accurately determine the distribution of this error. In Fig. 6 we show that 99 per cent of the time the error is less than ±1 per cent, and that 80 per cent of the time the error is less than ±0.5 per cent. One should note that since the error has a normal distribution with the mean at zero, the statement that 80 per cent of the time the error is less than 0.5 per cent is the same as saying that the error is less than ±1 per cent for 99 per cent of the time.

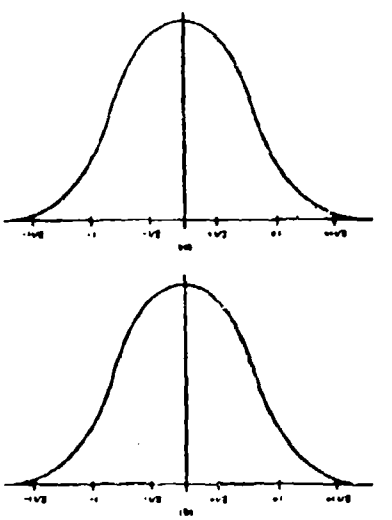


Figure 6 (a) Discriminator Linearity Error Distribution, 98% Area Lies Between ±1%; (b) 80% Area Lies Between ±1/2%.

(4) The error in the video response (Fig. 7c) also is defined as having a normal distribution with an error less than ±3 per cent occurring 99 per cent of the time.

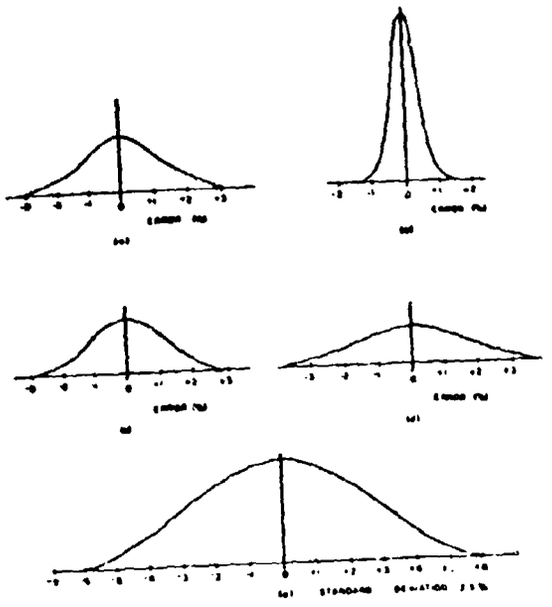


Figure 7 The Distribution of Errors in the Test System of Figure 1 (a,b,c, and d) Give a Total Error Distribution (e) for the System with Standard Deviation of 2.5%.

The distribution of the individual errors has been defined above. However, it is more convenient to express the distributions of the individual errors in terms of their mean and standard deviation (rms from the mean). From a table of the normal distribution it is observed that 90 per cent of the time observation will not exceed $\pm$ 1.65 standard deviations, and that 99 per cent of the time the observation will not exceed $\pm$ 2.58 standard deviations. Hence, the standard deviations of the individual errors, for e_i = 1,2,3,4 is determined by

$$\sigma_1 = \frac{2\%}{1.65} = 1.21\%$$

$$\sigma_2 = \frac{3\%}{1.65} = 1.82\%$$

$$\sigma_3 = \frac{1\%}{2.58} = 0.388\%$$

$$\sigma_4 = \frac{3\%}{2.58} = 1.16\%$$

The standard deviation of the test system, σ_T, is given by

$$\sigma_T^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2$$

$$\sigma_T^2 = 1.48 + 0.15 + 1.35 + 3.33$$

$$\sigma_T = 2.5\%$$

The mean of the system error is equal to the sum of the means of the individual errors. Since the mean of all individual errors is zero, the mean of the system error is zero. Hence, the test system in Fig. 1 has an error that is normally distributed with mean zero and standard deviation equal to 2.5 per cent. The system error is illustrated in Fig. 7e. Knowing the mean and standard deviation, the probability that the system error lies within given limits can be determined from a table of the normal distribution.

Let P_j per cent (j = 3,4,5,6,7,8,9) equal the probability that the system error is less than $\pm$ j per cent. Then, for the test system in Fig. 1

$$P_{3\%} = .7698$$

$$P_{4\%} = .8904$$

$$P_{5\%} = .9545$$

$$P_{6\%} = .9836$$

$$P_{7\%} = .9949$$

$$P_{8\%} = .9986$$

$$P_{9\%} = .9997$$

In the example, it is assumed that all the individual errors are distributed according to the normal distribution. This, as seen in Figs. 4,6, and 7, means that an error near the mean is most likely and that a larger error has less chance of occurring. In most cases, the errors have a normal distribution because there are many factors that can cause an error. The errors of each of these factors combine to yield an individual error that is approximately normally distributed.

The system error may be a combination of many individual errors that are not normally distributed. However, the system error is still approximately normally distributed with a mean equal to the sum of the means of the individual errors, and a variance (the square of the standard deviation) equal to the sum of the variances of the individual errors. This is true because of the central limit theorem.[1,2]

The following example is presented to illustrate how individual errors, which have different distributions, are combined. The individual errors are selected so that the system error is approximately normally distributed. Usually a system error that is composed of four individual errors having different distributions will not have as near the normal distribution as the system error in the example.

In the example,

(1) The first individual error is between $\pm$ 2 per cent. An error outside these limits is not possible, and an error of any value between these limits is equally likely.

(2) The second individual error is equally likely anywhere between zero and + 2 per cent.

(3) The third individual error occurs between zero and +4 per cent, and is most likely at +2 per cent. The probability of observing an error greater or less than + 2 per cent decreases linearly with the distance from + 2 per cent.

(4) The fourth individual error has a normal distribution with a mean equal to -1 per cent, and a standard deviation of 1.414 per cent.

The system error, which is a combination of these individual errors, is approximately normally distributed with a mean of +2 per cent, and a standard deviation of 2.09 per cent. (See Fig. 8.)

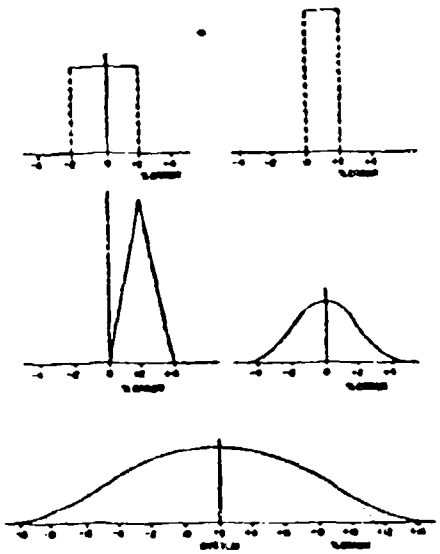


Figure 8 Error Distributions 1,2,3, and 4 Give a System Error Distribution, Approximately normally distributed.

These results of these examples show clearly that many test systems are more accurate than they appear to be under the summation of magnitudes calculation.

In the system illustrated, the system error was calculated to be ±10.5 per cent by the summation method; yet this same system has an error equal to or less than ±7 per cent for 99.49 per cent of the time.

If the test were run once a day on different equipment for one year, the system error would be greater than ±7 per cent only two times. If a ±7 per cent error were all that were required, it would be difficult to justify additional expense and time delays to increase the accuracy of the system.

Notice that the accumulating errors are less than ±3 per cent for 76.98 per cent of the time; yet this system yielded a ±10.5 per cent figure by the summation of magnitudes method.

Examine the consequences of a calculation by summing the positive and negative magnitudes: A telemetry system for satellite communications is designed. The performance characteristics of this system must be checked by a test system; one particular parameter must be checked, and it has a tolerance of ±5 per cent. A minimum measuring accuracy of 2.5 to 1 is insisted upon. A test system, having accumulating errors totaling ±4 per cent by the summation method, may be purchased. The system can be purchased for $11,000 on a six-week delivery schedule. However, to purchase such a system having an error of ±2 per cent, a six-month time delay is incurred and costs rise to $300,000.

The statistical calculation of the error in the original test system would have indicated a test system error of less than ±2 per cent for 98 per cent of the time.

The calculation based on statistical theory is not limited to errors in test systems. This technique can be used on any system or equipment having accumulating errors.

The assistance of Wray D. Dickens and John R. Pennell is appreciated. The assistance of John S. Hosford who edited and corrected the statistical theory was particularly helpful.

1. D. Frazor, Statistics, An Introduction, John Wiley & Sons, Inc., New York, 1958.

2. A. Mood, Introduction to the Theory of Statistics, McGraw-Hill, New York, 1950.

FACTORS INFLUENCING THE RESEARCH AND DEVELOPMENT OF NEW COMPUTER PROGRAMMING TECHNIQUES REQUIRED FOR MECHANIZATION OF MACHINE LEARNING

By: Dr. Robert E. Smith, Control Data Corporation

I. INTRODUCTION

Basic to the research being carried forward by Control Data Corporation in the general area of machine learning are the following premises:

(a) In order for a computer to approach the "machine learning" concept, it must be made capable of learning. This infers that the modern electronic computer must be given a tremendous increase in versatility capabilities over present repertories.

(b) In addition to the need for more versatility, is the desirability for reorganization of internal memory locations including new techniques of sequencing in order to more nearly approach the "human capability" concept.

(c) Finally, new techniques and expanded use of inter-communication between computer and man are required. Such techniques should develop conditions involving the constant flow of information between computer and operator (director of problem) so that adequate stimuli can be furnished to the computer for each problem situation.

This report is an abstract of some of the principles evolved from research involving these premises. It is further limited to the programming simulation phase of that research.

II. VERSATILITY

Early in the research it was decided to examine various programming areas which are generally concerned with repertoire versatility, in an effort to augment these capabilities. Since there is insufficient time to report on all areas that have been examined, the area of decision making has been chosen as representative of this part of the research.

Decision Making: Involved in learning, in such a way as to be almost inseparable, is the area of Decision Making. As an example of possible computer versatility expansion this area was selected, from the many others which were similarly investigated, as an example of how versatility can and must be extended if computers are to become capable of "learning"

What is involved in a person making a decision? The answer to this question is often obscured by the decision itself—in this case, the output. One tends to forget the "preliminaries"--the input--when faced with the "finalities". Careful analysis will indicate the latter is hardly possible without the former. Consider the effect of changing the decision maker in a problem situation. This is in essence the same as changing much of the input. Dramatic differences may show up in the decision itself by such a procedure. Even a slight change in the original input may affect the final result—for example, a new jury member during a trial, or a change of weather in a logistical plan of attack! Consider piecemeal, there is very little magic concerned with a human decision. The decision maker is presented with certain facts, stimuli and alternatives. Added to these are unique heridity and enviromental factors including his past experience, training, and education. Other subtle influences-perhaps prejudices and external pressures are included. All this input, rolled into a ball of some sort, is batted back and forth during the thinking process and a decision is made. If one were to hand the computer the same ball and provide an equivalent or similar "rule of solution" use by the human, would the computer arrive at a similar decision? One of the conclusions of this report is that only through increasing the computer versatility characteristics, so as to more nearly approach those contained in the human organism, is this possible.

Certainly such a provision is difficult in the area of decision making, if the decision-making power available is limited to plus-minus, or zero-non-zero capabilities. What are some of the decision-making characteristics which are desirable if the computer is to become a better thinking machine? In trying to find answers to this question, one is led to a consideration of factors which appear to influence decision making. These factors can probably be categorized by types as the following:

(1) Amount or Quantity	(2) Opposites
Size	Near-Far
Cost	Yes-No
Distance	Zero-Not Zero
Time	Plus-Minus
etc.	Fast-Slow
	Small-Large
	Hot-Cold
	East-West
	etc.

(3) Composites	(4) Sensual
Alternatives	Repetition
Practicality	Prejudices
Possibilities	Training
Probabilities	Experience
Efficiency	Instinct
Consequences	Practice
Similarities	Motivation
Trial & Error	Knowledge
Elimination	Health
Translation	Esteem and Welfare
etc.	etc.

Looking over this list one reaches the conclusion that decision-making on a computer generally involves the "opposites" category above, with little or no attention being paid to the others. Certainly part of the difficulty lies in the "either-or" characteristics of binary logic, yet in the final aspects of human decision-making, this "either-or" concept is basic. This leads one to conclude that the real difficulty lies not so much in the essential duality of logic as much as the fact that only these two characteristics are used. A somewhat analogous situation in humen decision-making might be the decision of an infant as to whether he should or should not eat.

To increase computer versatility in decision-making it is necessary to empower the computer with some of the conceptual abilities inherent in all other categories mentioned above. To sacrifice many of these for the benefit of one—for example, to build into the computer a statistical or probability formula and to base all decision making on this formula —is just as dangerous a position as to depend solely upon the "either-or" characteristics. What is needed is a general increase in versatility over all categories listed above so that the computer has a variety and range of capability extending over most problem solving situations.

What are some of the versatility characteristics a computer should have for improved learning capabilities? Before enumerating a partial list of suggestions in this area, it is important to note that any computer repertoire of instructions can be made more versatile by the following methods:

(a) Expansion of Present Repertoire

The extension and expansion of functions already performed by the regular computer instruction repertoire- for example, two-way jumps can be made into 3, 4, 5, ---n way jumps: consecutive search instructions can be made into non-consecutive or "skip-search" instructions; "store A at m" can be made into "store A at m_1, m_2, m_3——m_n", etc.

(b) Creation of New Instructions

New instructions can be devised either through interpretive combinations of regular instructions (by programming) or by actual hardware components. These new instructions can be particularly tailored to the learning process after a period of trial and error through programming simulation.

For this report, the suggestions which follow are aimed at the latter type rather than the former. They are also type instructions which the author feels can be expanded into several sub-instructions which would have particular impact on the learning process. (Arbitrary names are assigned to these new instructions for ease of reference.)

(a) RECALL INSTRUCTION PATTERN

Function: Recall Previous Use of Instruction(s).

This instruction will indicate the starting address(es) of the instruction(s) identical to the "n" instruction(s) starting at the given address "m".

(b) RECALL BIT PATTERN

Function: Recall Previous Use of Bit Pattern(s).

This instruction will indicate the starting address(es) of the bit pattern(s) identical to the "n" bit pattern(s) that are stored in consecutive memory locations starting at the given address, "m".

(c) INDICATE SIMILARITY COUNT

Function: Compares Similarity of Characteristics of Register Contents.

This instruction will indicate by count the degree of similarity of "C-characteristics" of the contents of n consecutive registers whose starting address is given with the contents of an equal number of consecutive registers whose starting address is in Q. (The "C-characteristic" can be a specific bit pattern, e.g. the operation codes, the operands, or the whole word.)

(d) EXECUTE THE NEXT SIMILAR INSTRUCTION

Function: After Next Instruction Jump to a Similar Instruction.

After the next instruction is executed the computer executes that instruction following which has a "C-characteristic" identical to the "C-characteristic" of the instruction just completed. The "C-characteristic" is a specific bit pattern, e.g. operation code, operand, etc.

(e) CONDITIONAL REPETITIVE EXECUTE OF NEXT INSTRUCTION

Function: Execute or Skip the Next Instruction Depending upon The Number of Previous Executions.

If the next instruction has been previously executed n times, skip; otherwise, executes.

(f) CONDITIONAL RELATIVE EXECUTE OF NEXT INSTRUCTION

Function: Execute or Skip the Next Instruction Depending Upon the Combined Bit Pattern of the Registers on Either Side of the Given Address.

This instruction examines the contents of the registers on either side of the register whose address is given and executes or does not execute the next instruction dependent upon the similarity of the "C-characteristic" of either (or both) of the bordering registers as compared to the "C-characteristic displayed in A.

(g) CONDITIONAL EXECUTE DEPENDENT UPON "N" LOOPS

Function: Execute or skip the Next Instruction Depending Upon Whether or Not the Routine Whose Starting Address is Given has been Repeated "n" Times.

This instruction examines the contents of the index register governing the number of loops made through a subroutine whose starting address is given, and on the basis of the number of loops indicated, executes or skips the next instruction.

(h) REVERSE SEQUENCE

Function: Reverse the Sequence of n Instructions Starting at Address L_1, and Ending with Address L_2.

This instruction reverses the execution of a sequence of instructions. The number of instructions, and the starting and ending addresses are given.

(i) MODIFY AFTER "N" MICROSECONDS

Function: Modify the Instruction at Address L_1, by the Logical Product of the Contents of Q and L_1, After n Microseconds.

This instruction will modify the contents of a designated register after n microseconds have elapsed from the start of the current instruction.

(j) RELATIVE CHANGE

Function: Change the "n" Registers on Either (or one) Side of the Register Whose Address is Given, by the Corresponding Bits in Q.

This instruction complements the corresponding bits of n registers on either (or one) side of the given address by the corresponding ones (or zeros) in Q.

(k) SEQUENCE

Function: Execute "n" Instructions by Sequence of Given Addresses and Return.

This instruction will execute n instructions in the same sequence as the given addresses and return to the instructions following the last given address.

(l) INVERSE

Function: Invert n Instructions Located at Given Address

This instruction will invert specific instructions located at given addresses and return to the next instruction following the last given address. (Certain instructions can be inverted, e.g. LOAD A↔STORE A, ADD↔SUBTRACT, MULTIPLY↔DIVIDE, INCREASE↔DECREASE REPLACE ADD↔REPLACE SUBTRACT, SHIFT LEFT↔SHIFT RIGHT, etc.)

(m) CONTINUE TO FIRST "K" INSTRUCTION

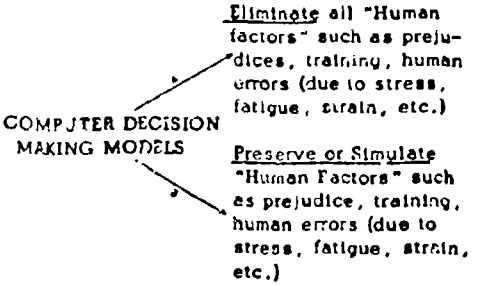
Function: Start with the instruction whose address is given and execute in sequence until a similar instruction to the one given is reached, then return.

This instruction will execute, in sequence, the instructions starting with the given address until an instruction similar to the given instruction is found. Then control is returned to the instruction following this order. ("Similar instructions" can have same operation codes, etc.)

Factors Influencing Possible Models

Time does not suffice to go into details of the programming simulation models being constructed or contemplated in the various areas of machine learning. Nevertheless, a few descriptive suggestions should indicate the possibilities and opportunities for such efforts. Again decision making is selected as a representative area and factors influencing possible models are briefly outlined in this area.

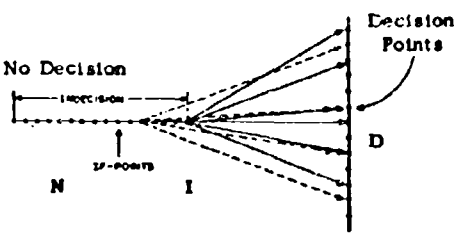
(a) As previously mentioned, human decision-making often involves much of the "human element". Ones training, background, prejudices, etc. often affect his decision. In computer decision making it may be desirable to preserve this human element (simulate it) or to eliminate it entirely from the model. Two different models are possible.



If B is selected, the means of preservation or simulation of "human factors" becomes important. Two alternatives appear to be possible. One, and the more difficult, is to build into the model, rules or statistical formulas which will represent "general" or "expected" weighting factors which will approximate the "human factors" prevalent for the "average" decision maker. The second, which at the same time suggests a whole new area for computers, is to let the computer actually measure to extent or degree to which these human factors are present in the decision maker. It is hard to imagine a more proficient machine than a computer by which to measure how well a person can react to stress, fatigue, strain, prejudice, and many other subtle factors which often affect human decisions. The gained computer grade can then be used in the remaining mathematical algorithm to weight the decision toward one or more of several possible paths.

(b) Many human decisions are based upon an "alternative-consequence" concept. One "looks into the future" along alternate paths, wherein consecutive events take place leading toward certain consequences at the end of each road. It is in this area in particular that the computer may be made to surpass the human decision making. Whereas the human is apt to see only these alternative paths and events which he "wants to see" due to lack of time, personal estimation of what is or is not important, difficulty of having too many things to think about, etc.) the computer can be made to spill out the consequences of hundreds of alternative events for examination. One method of making the model possible is to assign each event an "accomplishment" percentage value of each consequence. All of the different possible combinations of events can thus be compared along with the relative values representing the consequences.

(c) Instead of reaching a decision, one generally approaches a decision. Many problem situations are analogous to the decision model shown below:



There is a period of "no decision" when one is examining, investigating, and weighing involved parameters. This is also a period of indecision when one is not sure of the road ahead. Nevertheless, one is closer to reaching a decision at this point (I) than at the starting point (N). If the point I is moved closer to the point N, it is entirely possible that a different pattern of radial paths toward some

decision may result. Thus a "hasty" decision may be different than a more "leisurely" one.

Each of the points along line NI can represent "If-Points". Each If-Point (when reached) weights the various possible radial decision paths by some amount. The sum of all "if-points" preceding a given "if-point" also weights the various radial decision paths.

Random number generators for the various if-point values and corresponding radial decision paths makes this model particularly interesting to simulate. Under actual model conditions the decision maker(s) can provide if-point values as input data dependent upon previous output from the computer.

III. REORGANIZATION OF INTERNAL STORAGE

Increased versatility in instructional power is not the only characteristic required to improve computer learning capabilities. Of equal importance is the reorganization of internal storage or memory locations and internal time sequences.

Many have claimed that a computer can never approach the human brain in learning because of the almost limitless memory capacities of the brain. One wonders whether it is capacity or the ability to "over-ride" and "store in layers" which is more important. Certainly it is evident that modern computers do not make full use of the memory capacity that has been built into them. How many computer programs utilize the complete internal storage capacities of the machine? The problem is further aggravated by the fact that computers are generally limited to one program at a time with a minimum of time-sharing activities being permitted. On the other hand, there seems to be some evidence, psychologically and psychiatrically, that the brain is capable of time sharing several series of thought processes in various states of ascendency.

Finally, there is the problem of internal sequencing of events. Here the computer generally calls up one instruction at a time and the sequential execution of the instructions that follow make up the "event(s)". One might argue that the brain calls up more than one thought at a time, peruses these, either leisurely or almost instantaneously, makes a choice, and the choice in turn then calls up a series of events which may lead to more selections, series of events, etc.

It would appear that the following internal storage and sequencing characteristics would improve the learning capabilities of a computer.

(a) The ability to saturate internal memory locations to their full capacities.

(b) The ability to time-share activities and some series of instructions.

(c) The ability to call up several series of instructions.

Again, the partial list which follows is suggestive of the methods that are being used to approach the above conditions.

1. Internal memory location saturation are being examined through two techniques:

   (a) By providing groups of memory cells with different capacities.

   (b) By arbitrary division of larger memory cells into smaller memory cells.

2. Time-sharing of events has been made possible to some extent through two techniques:

   (a) By exploitation of input/output mechanisms and an on-line smaller computer.

   (b) By the use of more than one current program address register.

3. Some success has resulted in simulating a series of events by a sequence of instructions whose location is triggered by a word, a special instruction, or a specific bit pattern of a word that is examined.

IV. NEW COMMUNICATION LINKS

Increased computer versatility and reorganization of internal memory are two important components required for computer learning capability. There is another needed characteristic that cannot be over-emphasized. This is wrapped up in the ability of the machine to communicate with something outside of itself. Take away all human sense organs and many of the stimuli for initiating thought processes are gone. Much human learning progress appears to be charted, routed, extended, or expanded by external stimuli which enter the brain from outside. One might argue, in fact, that an external device (e.g., a book, pencil, the spoken word) becomes a crutch enabling the brain to limp from one series of thoughts to another. There is a continual inter-flow of

stimuli and responses, by way of the senses, between the brain and this learning device. Surprising is the realization that this flow of recognition response signals is a reasonably slow process--it rarely takes place with the speed of light.

In the desire to gain greater speeds in computer input/output devices, the full exploitation of communication between computer and external equipment has often been neglected. In the area of problem solving or decision making the initial problem parameters often change or are modified before a conclusion can be attained. These new parameters which must enter the algorithm will of necessity influence or change the final outcome. Conceivably such entrance of new data should be possible while the problem is being solved. Even more intriguing is the problem whereby succeeding events are controlled by periodic input to the computer. This input in turn can be made a function of the output from previous executions. In such a situation the input/output media is an integral part of the simulated learning process.

Mechanized machine learning may always be a concept invented by man and never fully achieved by mechanical means. Certainly it is to man's best interests to always control the process. The belief, expressed in this report, and being developed by research at Control Data Corporation, is that machine learning can be brought closer to full control by improving machine capabilities in certain areas. Three of these areas are:

Repertoire Versatility
Internal Memory Organization and Time Sequences
Greater use of Input/Output Media

PATTERN RECOGNITION

By: J. W. ..ouillette and C. W. Johnson, Electronics Laboratory, General Electric Company

The development of geometric pattern recognition equipment has been receiving an increasing amount of attention in the last five years. As a result, machines for recognition of typed or printed characters using magnetic or optical scanning devices are becoming commercially available, and it is natural now to give more serious attention to problems involving identification of geometric patterns which display a greater range of variation in character than is normally the case for stylized, typed letters of the alphabet. In the fall of 19..3, the General Electric Company began a survey of the problems which will arise in the processing of optical data in the form of photographic reproductions of terrain, or photographs of other patterns having a geometric origin, but less standardized in nature than printed figures.

It was apparent at the onset of the investigation that the initial registration or alignment process can be a very time-consuming operation in any system directed toward analysis of the randomly positioned geometric shapes which are encountered in analysis of reconnaissance photographs or other forms of similar data. This, of course, is the process of bringing the stored reference pattern into some standardized orientation and horizontal-vertical relation with the unknown patterns in the scanner field, and is carried out either through motions of the scanning raster, or actual physical motions of the unknown pattern.

The registration phase as it occurs in typed character recognition systems assumes a less formidable role, since a single orientation alignment can usually eliminate angular misregistration for at least one line of type and frequently for a complete page, and a single vertical alignment is usually sufficient for a complete line. However, in the problem of reconnaissance photograph analysis, for example, there is no likelihood that objects of interest will always be encountered in any standardized relation in the scan field. This applies to orientation, horizontal and vertical position, magnification and even perspective. Alignment through electrical or physical orientations and positionings of a raster or stored reference is, of course, still possible; however, the large number of combinations which must be covered, when four or more parameters must be varied to fix exact registration, means that alignment becomes a very time-consuming part of the overall operation, particularly when it must be repeated for each pattern of interest.

An alternative to actual raster alignments is possible, however, and this was suggested in broad terms many years ago by those studying the pattern function of the human brain. This approach depends on detection of intrinsic or "Gestalt" features which are independent or invariant with respect to misalignment. The General Electric Company instituted a study and development program which showed through equipment demonstration that pattern recognition based on this approach is feasible, and may be a preferable approach when the number of transformations causing misregistration is large.

The concept of geometric invariance is well-known to mathematicians, but the manner in which the idea can be used in identification problems has not been particularly emphasized. The function known as "radius of curvature" is a good example of a function invariant under vertical, horizontal and angular displacements, and it is useful for identification, since it is unique. That is, two curves which possess the same radius of curvature plot will always be congruent. In contrast, other functions, for example the enclosed area of a closed curve, are also invariant; however, these functions do not uniquely identify the pattern in question, since other patterns may possess the same invariant. Figure 1 illustrates the relationship of a plot of the reciprocal of radius of curvature to the specimen pattern shown in Figure 1A. This pattern is made up of semicircles of three different radii, and yields an inverse plot of $K = 1/R$ as shown in Figure 1B. (Negative and positive values of K are used according to a convention based on the direction of travel in traversing the original curve.) In Figure 1C, the original curve has been rotated and translated, but, of course, it still yields the same plot for K. The problem now has been changed to recognizing the plots of K through some correlation procedure rather than the plots of the original pattern.

Now there remains a certain ambiguity in the horizontal positioning of the K plots, which stems from the lack of a uniquely determinable reference point on the original curve. Such a reference point would be used as an origin from which to start the tracing of the original curve and the plot of K. Without a uniquely defined origin, a horizontal registration search is still required before a correlation process can identify the K plot. However, a rather simple procedure can be employed which will eliminate even this single search for registration. This consists simply in plotting the invariant K, not against position on the original curve, but against another independent invariant. The results of this procedure for a general case are illustrated in Figure 2. Here the original pattern, shown in Figure 2A, is transformed into a new pattern in Figure 2B, with the plot against a new pair of axes or new functions I_1 and I_2 which are invariant under rigid motion and magnification. Since the functions I_1 and I_2 are invariant under these transformations, the second pattern shown in Figure 2C, which is derived from the pattern just above through a combination of these transformations, yields in the I_1 and I_2 plane a pattern which is

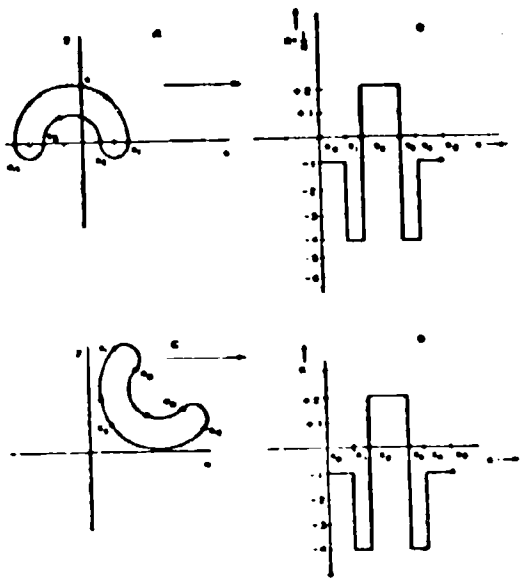


Figure 1

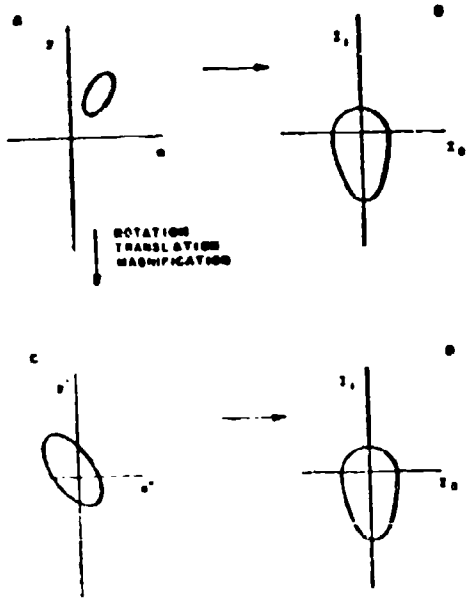


Figure 2

identical to the pattern shown in Figure 2B just above. Thus, for identification of patterns of the type in Figure 2A or 2C, we need only design our equipment to identify the single pattern of the type in Figure 2B and 2D, and now no registration is required.

The discovery of functions which are invariant under rigid motion and magnification is not particularly difficult; however, proofs that these functions provide unique characterization is not always so straightforward. In addition, some discretion must be exercised in selection, since problems associated with electrical implementation and noise immunity must be kept in mind. For example, the function dR/ds, where R is radius of curvature and s is arc length, is invariant under rigid motion and size changes, but the presence of third derivatives in this function means that judicious filtering should be used to accentuate the overall pattern and to suppress frequency components in the curvature function resulting from local irregularities in the pattern which do not contribute to the identification.

IMPLEMENTATION

A system for laboratory demonstration of the invariance concept was completed in 1956. Since the basic aim of this system was to dramatize the fundamental concepts, many of the basic functions were implemented in a manner designed specifically for this goal, rather than for a general purpose field equipment. With these considerations in mind, the initial scanning operation was based on a high speed, constant speed electronic curve follower, rather than a more conventional rectangular scan raster. The follower used for the demonstration system was based on a design which very conveniently permitted generation of invariant functions. In addition, since constant speed followers are not particularly well-known, some detailed information concerning the design will be worth-while, although the follower was not the primary goal of the study.

The data to be handled by this follower were prepared as black-on-white line drawings, from which contrast photographic negatives were prepared. These negatives were placed in an optical train consisting of an electrostatic deflection cathode-ray tube, a lens system imaging the face of the tube on the negative, a negative carrier, a collecting lens, and an end window multiplier photo-tube. A schematic of this arrangement is shown in Figure 3.

In order that the flying-spot scanner shall trace out the curve, it is necessary to derive information about the curve direction at each point. This was done by causing the spot on the tube to move in a small circle at a high rate.



Figure 3

This circle will be referred to as a "search circle". This was accomplished in the usual manner by applying small sinusoidal signals of equal amplitude to the x and y deflection systems, differing in phase by 90 degrees. The frequency in our apparatus was 450 k cps, which allows the use of ordinary broadcast receiver components.

If the search circle is so located that it intercepts the curve each time the image of the flying spot crosses the curve, the photocell will be illuminated, and a pulse of current will flow in its load. These pulses were amplified in a broad-band video amplifier to preserve rise-time and clipped rather severely. The result of this operation was a train of pulses of uniform amplitude and width, whose time occurrence can be shown to contain information both as to the distance from the center of the search circle to the curve and also as to the direction of the curve. The train of pulses was applied to two filters as seen in Figure 4. One of the filters was tuned to the search frequency and the second was tuned to the search frequency second harmonic. It can be shown that if the circle were centered upon the curve, the fundamental content would be zero and the second harmonic would be a maximum. Furthermore, if the center of the circle were slightly displaced from the line, then the fundamental output would increase linearly, changing phase by 180 degrees as the circle center crossed the line. The second harmonic would go through a broad maximum with no phase shift.

The second harmonic output was applied as a synchronizing signal to a slave oscillator tuned to the fundamental search frequency. The output of this oscillator was then a carrier of the same frequency as the search circle. It can be shown that the phase angles between this signal and the search signals are numerically identical to the actual geometric angle between the direction of the line and the axes defined by the x and y deflection systems.

The signal from the fundamental filter, having a phase reversal at the center, is a measure of the error in circle displacement. It is either in phase or 180 degrees out of phase with the synchronized oscillator output. This signal was shifted by 90 degrees in an RC phase shifter. The shifted signal was then added linearly in a resistor network to the synchronized oscillator signal. The result of this operation was applied to an AGC amplifier. The output of this section was then a carrier of constant amplitude. If the search circle is centered on the curve being followed, the phase difference between this signal and the x-axis search signal is equal to the slope of the curve at this point. If the circle is displaced slightly, then the phase difference is increased or decreased, depending on the direction of the displacement.

This constant-amplitude signal was then treated as an analogue vector velocity. The constant amplitude implies constant velocity magnitude,

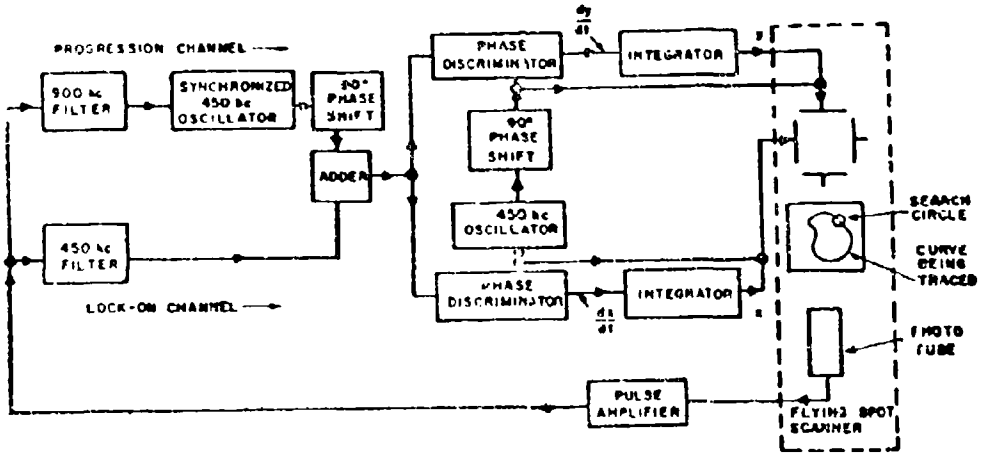


Figure 4

or constant speed. The phase is the direction. A pair of phase detectors were used to derive the x and y components of velocity. One phase detector had as reference the carrier supplying the x search, the other had the y search as reference.

The voltages representing rectangular components of velocity were integrated once, to obtain analogues of the x and y coordinates of a spot moving at constant speed along the curve. The voltages representing coordinates were added to the search voltages and the resultant signals were used to control the motion of the spot.

The constant speed of motion of the spot makes a linear relation between time and distance measured along the curve from any arbitrary reference. Any invariant function of arc length can be computed from the voltages representing velocity, direction, and position, as the beam traces out the curve in a repetitive manner.

As an example of the calculation of an invariant, we shall consider the curvature. This is defined as $d\theta/ds$, where θ is the direction angle of a curve and ds is the increment in arc length. We write $d\theta/ds = d\theta/dt \cdot dt/ds = 1/u \cdot d\theta/dt$; where u is the speed, or magnitude of the velocity. We recall that in the curve-tracer, u, the speed is held constant. Thus the problem is solved if we calculate $d\theta/dt$.

We note that in the curve-tracer, there is a carrier of nominal frequency equal to the search frequency, with varying phase where the phase is numerically the direction angle of the curve. This voltage can be expressed as $E_1 \cos(\omega t + \phi)$, where E_1 is a fixed amplitude, ω and t have the usual connotation, and ϕ is, as usual the varying phase. ϕ is, of course, numerically identical with θ, the direction angle. We expand ϕ in a Taylor series -

$$\phi = \phi_0 + \frac{d\phi}{dt} t + \cdots$$

If ϕ varies slowly, then we ignore higher terms and write $\phi = \phi_0 + t\, d\phi/dt$. Then the velocity analogue voltage can be written as $E_1 \cos(\omega t + \phi_0 + t\, d\phi/dt)$; or $E \cos\left[(\omega + d\phi/dt)\, t + \phi_0\right]$.

This implies that a continuous change in direction evinces itself as a frequency shift in the velocity analogue voltage. We can calculate $d\phi/dt$ in a very simple manner. An FM discriminator tuned to the search frequency has the constant amplitude velocity analogue voltage as input. For deviations within the linear range of the detector, the output of the discriminator will be proportional to the rate of change of phase with time, and hence to the rate of change of direction with time, and finally, since the tracing speed is constant, directly proportional to the radius of curvature, which is the desired invariant function.

Other invariant functions are also easily generated by rather simple processing of various voltages generated in the curve follower. For example, the invariant function defined by the angle formed by the radius vector to a running point on the curve and the tangent to the curve at that point is clearly independent of not only all rigid motions, but also magnifications. This construction is sketched in Figure 5. Generation of the sine or cosine of this angle can be carried out by the balanced modulator, adder and phase discriminator circuit shown in Figure 6. The condenser-resistor combination shown as the plates of the CRT serve to locate the centroid of the curve and thus the origin for the radius vector. The final step in the recognition process then consists in cross-correlation of the invariant plots with standard plots in the system storage.

As suggested before, specific embodiments of the general mathematical principle of geometric invariance are not uniquely tied to the analogue curve follower system described, but the principle itself is fundamental and almost always must be used in any system in which geometric patterns are to be correlated with a minimum of trial and error in registration and alignment.

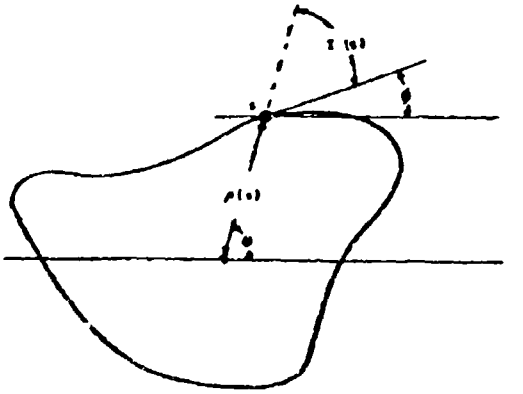


Figure 5

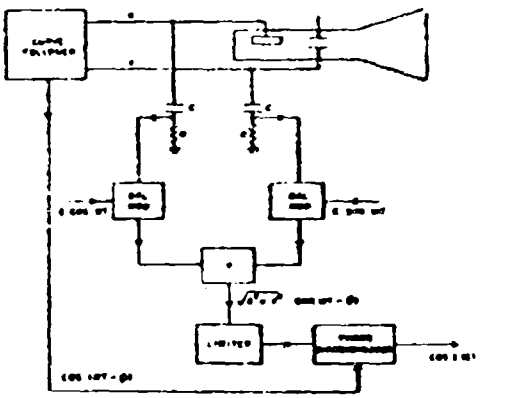


Figure 6

NEW TECHNIQUES IN RESIDUAL ARITHMETIC

By: Messrs. M. Levine, J. Marx, S. Levy; Arma Division, American Bosch Arma Corporation

Abstract

The paper discloses the application of residual numbers to arithmetic processes in ultra high speed computing machines. Methods of numerical manipulation for division, multiplication and rounding, relative magnitude determination, fractions and polynomial evaluation are suggested so that the inherent virtues of residual arithmetic techniques are available to the machine designer.

Introduction

This paper describes the use of a composite modular arithmetic number system to be called the Residual Number System for Computers.[1] The system, in operation, requires no carries between digits and so allows single step, parallel, arithmetic operations for addition, subtraction and multiplication. The paper also discusses the problems associated with the use of residual numbers for computation, and offers, in detail, a set of solutions to these problems that may make it feasible to build a general purpose, ultra high speed computer using these techniques.

Definition of Least Positive Residues

Let m be an integer, and consider the following linear equation which defines the residue R, where N, k, R are integers:

$$N = k\,m + R$$

Then N is said to be congruent to R, modulo m. An equivalent definition is that N is congruent to R modulo m if and only if m divides N - R. This relation is denoted by $N \equiv R \pmod{m}$.

If $S = (r_1, r_2 \ldots r_m)$ is a system of m integers, then S is a complete system of residues modulo m if and only if the r_i are incongruent, pair-wise, modulo m. In particular, S = (0, 1, 2 ....., m-1) represents the complete system of least positive residues.

Using m = 3, the ordinary numbers can be written as:

N = k m + R
0 = 0 3 + 0
1 = 0 3 + 1
2 = 0 3 + 2
3 = 1 3 + 0
4 = 1 3 + 1
5 = 1 3 + 2
6 = 2 3 + 0
. . . . .
. . . . .
. . . . .

So that, for example:

The numbers 0, 3, 6.....3n $\equiv$ 0 (mod 3),
1, 4, 7.....(3n + 1) $\equiv$ 1(mod 3), etc.

For m = 7:

The numbers 0, 7, 14 .... 7n $\equiv$ 0 (mod 7),
1, 8, 15 ....(7n + 1) $\equiv$ 1 (mod 7),etc.

Arithmetic Properties of Residues

If m is any fixed integer, then the set of least positive residues modulo m form a commutative ring with unity, in that the arithmetic operations of addition, subtraction and multiplication are well defined.

These arithmetic operations can be illustrated by the following tables:

addition mod 3

	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

multiplication mod 3

	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Considering our fundamental relationship,

$$N = km + R,$$

then $8 = 2 \cdot 3 + 2$
$20 = 6 \cdot 3 + 2$

or rewriting

$8 \equiv 2 \bmod 3$ and $20 \equiv 2 \bmod 3$

Evaluating (20 + 8) 8 in the module 3 system yields

$((2 + 2)\,2) \bmod 3 \equiv ((1)\,2) \bmod 3 \equiv (2) \bmod 3,$

or

$$(20 + 8)8 = 224 \equiv (2) \bmod 3$$

Let $m = (10)^6$ and form a one-to-one correspondence between the non-negative integers less than 10^6 and the least positive residues mod m.

Then,

$$N = k\,m + R$$
$$0 = 0\ 10^6 + 0$$
$$1 = 0\ 10^6 + 1$$
$$2 = 0\ 10^6 + 2$$
$$\vdots$$
$$m-1 = 0\ 10^6 + m-1$$

We can add, subtract, and multiply the residues, preserving correspondence, providing that the result, y, of an arithmetic operation falls between the limits 0 and M; i.e., $0 \leq y \leq m$.

If this were done, residue operations would be exactly the same as integer arithmetic operations.

Suppose we wished to do these operations in one step. One approach would be to set up a $(10^5)^2$ decimal magnetic core matrix for arithmetic evaluation. Another approach would be to use the Residual Number System which is essentially a method of dividing up the (10^5) matrix into approximately ten 100-element matrices which when interrogated, in parallel, yield the arithmetic result in residual notation.

Definition of Residual Number System

Let $m_1, m_2, \ldots m_n$ be relatively prime moduli and consider for each m_i, $i = 1, 2, \ldots n$, the least positive residues mod m_i. Let $M = \prod_{i=1}^{n} m_i$

Then given an arbitrary integer N such that $0 \leq N < M$, there exist n unique integers r_i such that $N \equiv r_i \pmod{m_i}$, and conversely, given a set of n elements r_i mod m_i, there exists one and only one integer, N, in the interval $0 \leq N < M$ corresponding to the r_i.

Arithmetic Operations of Numbers Encoded in the Residue Number System

Let

$$A = a_1 \bmod m_1, a_2 \bmod m_2 \ldots, a_n \bmod m_n$$

and

$$B = b_1 \bmod m_1, b_2 \bmod m_2 \ldots, b_n \bmod m_n.$$

Then

$A + B = C = (c_1, c_2, \ldots, c_n)$ where

$c_i = a_i + b_i \pmod{m_i}$, and

$A \cdot B = C = (c_1 \ldots, c_n)$ where

$c_i = a_i \cdot b_i \pmod{m_i}$, and

$A-B = C = (c_1, \ldots, c_n)$ where

$c_i = a_i + b_i$.

$A/B = C = (c_1 \ldots c_n)$ where

$c_i = a_i \cdot b_i^{-1} \bmod m_i$; $b_i^{-1} \cdot b_i = \bmod m_i$

This division only exists if all b_i are not equal to 0 and A/B is an integer. It should be noted at this point that the arithmetic is uniquely defined only when $0 \leq C < M$ and that the C_i are reduced to least residues mod m_i.

Suppose we wished to subtract integer $N_1 < M$ from a smaller number N_2, then

$$N_2 - N_1 = N_3 < 0$$

Since there exists only a one-to-one correspondence between positive integers and the residue system, negative numbers are not defined. Therefore N_3 is not defined and thus arises a need to define negative numbers.

Subtraction and the Representation of Negative Numbers

Consider this set of M ordinary integers:

$$M = 0, 1 \ldots \frac{M-1}{2}, \frac{M}{2}, -\left(\frac{M}{2}-1\right), -\left(\frac{M}{2}-2\right) \ldots, -1) \text{ and}$$

place them in one-to-one correspondence with a residue number system consisting of M unique states. Then it is clear that any integer $N + \bar{N}$ (its inverse) is congruent to 0 mod m_i and that addition, subtraction and multiplication of x and y are defined provided that $x \begin{pmatrix} + \\ - \end{pmatrix} y < \left|\frac{M}{2}\right|$.

For example:

Let $M = 11$, $m_1 = 2$ and $m_2 = 7$, then there exists a one-to-one correspondence between the set of integers $N < \frac{M}{2}$ and the 11 unique residue states.

Decimal	Modular Mod 2	Modular Mod 7	Decimal	Modular Mod 2	Modular Mod 7
0	0	0	-6	0	1
1	1	1	-5	1	2
2	0	2	-4	0	3
3	1	3	-3	1	4
4	0	4	-2	0	5
5	1	5	-1	1	6
6	0	6			
7	1	0			

and observe that

$$6 + \bar{6} = 0, \quad 6 + 0, \quad 1 \neq 0, \quad 0 = 0$$

and

$$-2 \times -2 = 4 = 0, \quad 5 \times 0, \quad 5 \neq 0, \quad 4.$$

In other words, all residue representations of integers greater than M/2 are considered negative while all numbers less than M/2 are considered positive. This process is completely analogous to using 2's complement arithmetic in conventional binary computers.

Conversion of Residue Code to a Mixed Radix Number

In order to interpret residual numbers, a conversion scheme must be employed to convert residual numbers into a positional notation such as binary, decimal or mixed radix numbers.

The mathematical justification for conversion into a mixed radix code is based on the following theorem:

Given a set of relatively prime moduli, a necessary and sufficient condition for a residual number N to be divisible by one of its moduli is that a zero exists in the column that corresponds to the modulus that is used for division. Division by a modulus can be executed with single-impulse circuitry by multiplying the residual number by a stored multiplicative inverse of the modulus.

Conversion Algorithm:

Let A be an integer encoded a_1 mod m_1, a_2 mod m_2 ...a_n mod$_n$, and let $k_n = a_n$, and let B = $(A + \overline{K}_n)$ mod m_1. Then B = $a_1 + \overline{k}_n$, $a_2 + \overline{a}_n \ldots a_n + \overline{a}_n$) which equals $(b_1, b_2, \ldots b_{n-1}, 0)$. Now B is divisible by m_n because $b_n = 0$. Dividing B by m_n yields an integer C defined mod m_i, i = 1, 2, ...n-1. If we let $k_{n-1} = C_{n-1}$, and K_n mod m_1 to C, and then divide C by m_{n-1}, we create a new integer D. Continuing this process we develop a sequence of integers k_n, $k_{n-1}, \ldots k_1$ which are the coefficients of a mixed radix number, with k_1 as the most significant digit. To illustrate, let the decimal number 65 be encoded in residual notation. Then the algorithm for conversion proceeds as follows:

Let $m_1 = 2$, $m_2 = 3$, $m_3 = 5$, $m_4 = 7$.

		Moduli 2	3	5	7	
The Decimal No.	65	1	2	0	(2),	therefore $k_4 = a_4 = 2$.
Adding $\overline{2}$	$\overline{-2}$	0	1	2	5	
	63	1	0	3	0	Since a zero appears mod 7, the number is divisible by 7.
(Multiplying by 1/7	1/7	1	1	3	1	
yields	9	≡	1	0	(4) X	$\therefore k_3 = 4$
Adding $\overline{K}_3$	$\overline{-4}$	≡	0	2	1 X	
yields	5	≡	1	2	0	resulting in a zero mod 5
Multiplying	1/5	≡	1	3	X	
yields	1	≡	1	(1)		$\therefore k_2 = 1$.
Adding $\overline{k}_2$	$\overline{-1}$	≡	1	2		
	1					
yields	0	≡	0	0		
Multiplying	1/3	≡	1	X		
yields	0	=	(0)			$\therefore k_1 = 0$.

Writing down the k_i digits in order yields (from right to left)

65 = 0 1 4 2,

or 65 = (0) + (3 x 5 x 7) + (1) (5 x 7) + 4 (7) + 2,

or 65 = 0 X 105 + 1 X 35 + 4 x 7 + 2,

where 105, 35, 7, 1 are the mixed radices.

Uses of Mixed Radix Code

Two things should be noted in the section that deals with subtraction and the representation of negative numbers: first, that the resulting sign of any arithmetic operation is implicit in the representation, and second, that all numbers greater than $\frac{M}{2}$ are considered negative. All that it is necessary to do in order to decide whether $a > b$ is to note whether a-b is greater or less than $M/2$. The decision can be made by examining the most significant digit of the mixed radix representation and determining whether it is greater than or less than $\frac{m_1}{2}$.

Suppose x and y are modular integers $< M$; then their product, $x \cdot y$, may be greater than M but must be less than M^2. In order to prevent the product from exceeding the number of defined states, let us adjoin an additional set of relatively prime moduli whose product is greater than M. Then the procedure for multiplication is as follows: Convert x and y to mixed radix code. Simultaneously consider each of the ordered code digits as they appear and encode them, mod the adjoined moduli, and then multiply the double precision numbers.

To illustrate, consider the product of the two decimal numbers x = 800 and y = 801 encoded mod 31 and mod 32.

		mod. 31	mod. 32
800	≡	25	0
801	≡	26	1

Converting x to mixed radix code

		Mod 31		Mod 32
800		25		0
0		0		0
800		25		0
X 1		X 1		X
32				
25		25		
800	= 25	(32)	+	0
800	= 25	0		

Converting y to mixed radix code

801		26		1
-1		30		31
800		25		0
X 1		X 1		X
32				
25		25		
801	= 25	(32)	+	1
801	= 25	1		

Now that x and y are encoded in mixed radix form to encode them mod the adjoined moduli 3, 7, 11 requires n - 1 parallel additions, because x = 800 = 25 (32) + 0 where the (32) is understood and is = 4 mod 7 (x) mod 7 = (800) mod 7

= 25 mod 7 x 32 mod 7 + 0 mod 7

= 25 mod 7 x 4 mod 7 + 0 mod 7

= 4 mod 7 x 4 mod 7 + 0 mod 7

= 2 mod 7 + 0 mod 7

= 2 mod 7

25 x 32 mod 7 can be inputed to the mod 7

adder without time delay. The same type of simultaneous computation will yield x and y mod 7, 11 and 13.

We proceed to multiply the double precision numbers and obtain a double precision product.

		31	32	::	11	7
800	=	25	0	7	8	2
X 801	=	26	1	8	9	3
640,800	=	30	0	4	6	6

Roundoff:

Fixed radix roundoff can be considered as the following linear function.

(1) $P = Q\,10^n + R$

where $P = a_1 10^{2n-1} + a_2 10^{2n-2} \ldots a_n 10^n \ldots$

$\ldots a_{2n} 10^0$ and is the number to be rounded,

$Q = a_1 10^{2n-n-1} + a_2 10^{2n-n-2} + \ldots a_{n-1} 10^{n-n}$

and $R = a_n 10^{n-1} + a_{n+2} 10^{n-2} \ldots a_{2n} 10^0$. If $R = \frac{10^n}{2}$,

then P rounded n digits = Q + 1; if $R \geq \frac{10^n}{2}$,

P rounded n digits is equal to Q.

Consider the following linear equations:

$$P = Q^1 (10^n + \epsilon) + R^1$$

then $$P = Q^1 (10^n) + Q^1 \epsilon + R^1$$

Clearly R^1 and Q^1 are less than $10^n + \epsilon$, and ϵQ^1 is less than $\epsilon(10^n + \epsilon)$, therefore $R^1 + Q^1 \epsilon < 1 + \epsilon(10 + \epsilon)$. Now $R^1 + \epsilon Q^1$ can be expressed as $(R^1 + \epsilon Q^1) = Q^{11} 10^n + R$ where R is the same R as in equation (1). Therefore P rounded n digits = $Q^1 + Q^{11} + 1$ if $R \geq \frac{10^n}{2}$ or $Q^1 + Q^{11}$ if $R < \frac{10^n}{2}$.

In the given example, it should be noted that the product of the adjoined moduli equals 7 X 11 X 13 = 1001 = $10^3 + 1$. Consider the following divisions

$$1001 \overline{)\,640800}\quad \begin{array}{r} 640 \\ \hline 6006 \\ 1020 \\ 1001 \\ \hline 19 \end{array}$$

where $Q^1 = 640$, $\epsilon = 1$, $R^1 = 160$
now $P = 640(10^3) + 800 - 640 + 1 = 641$ since $R \leq 500$ now P also equals $P = 640\,(10^3+1) + 160$
$P = 640\,(10^3) + 640\,(1) + 160$

Therefore to round P, simply add 160, the mixed radix remainder obtained after dividing by 1001 to $\frac{640640}{1001}$ mod. 31, 32 and decide in which interval R falls.

Polynomial Evaluation

Using a modular representation, the ability to evaluate polynomials with a priori determined coefficients in a single arithmetic operation stems from two observations of their structure in residual number representation.

1. If x is a modular integer (mod q), then:

$$x^n \pmod q = x^n \pmod{q-1} \pmod q.$$

This statement is a direct extension of Fermat's Theorem which states:

$$x^p \bmod p = x \bmod p$$

2. If P(x) is an arbitrary polynomial of the form:

$$P(x) = a_0 + a_1x + a_2x^2 \ldots a_nx^n$$

then P(x) in modular notation is:

$$P(x) \bmod a = a_0 \bmod a + a_1x \bmod a + \ldots a_nx^n \bmod a$$

$$P(x) \bmod b = a_0 \bmod b + a_1x \bmod b + \ldots a_nx^n \bmod b$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$P(x) \bmod m = a_0 \bmod m + a_1x \bmod m + \ldots a_nx^n \bmod m$$

It can be noted at this point that there are only "a" arguments and solutions mod a and "b" arguments and solutions mod b, etc; so that a relatively small set of tables of solutions ($\vdots$ m_i) can be used for single step evaluation, whereas in ordinary arithmetic, the size of the tables = $M = \prod_{i=1}^{n} m_i$

Mathematical Justification for Fractional Arithmetic:

The modular representation for any integer $p < M = \prod_{i=1}^{n} m_i$, where the m_i are relatively prime moduli (with respect to each other), forms a commutative ring with unity, in that the operations of addition, subtraction and multiplication are always defined.

The law of commutativity states that: if b and c are any arbitrary modular numbers then: bpc = pbc = cbp gives the same modular results regardless of the order in which the multipliers are taken.

Suppose then that b and c are chosen so that bc = 1; then bpc = pbc = cbp = p. This can be true if, and only if, $c = b^{-1}$; that is, c is the multiplicative inverse of b. The criteria for c to exist is that b must be relatively prime to each of the moduli. If b and c exist, then b^2, b^3, $\ldots b^n$ exist, and c^2, c^3, c^4, $\ldots c^n$ also exist; and $c^nb^n = 1$ for all n.

Now let $R^n = b^n$ and $R^{-n} = c^n$, where R is the radix of the integers used to establish the modular representation. Then the following fractional arithmetic can be defined:

Given p as any modular coded integer M:

$b^{2n}(c^np + c^npc^np) = b^np + p^2$. This follows, provided that $b^n + p^2 < M$, and the expression is interpretable as a unique integer.

Fractions

Suppose we wished to form the number .05. This number can be considered as 5 times a number x, where x is defined by the following relationships:

$$10^2 x = 1, \quad \therefore x = 10^{-2}.$$

i.e., x is the multiplicative inverse of 10^2. If we set up a one-to-one correspondence between the real integers, N, and a modular system, M, and if 10 is a real integer encoded in a modular system containing no zeros in any one of its modular digits, then 10^2 contains no zeros and, in general, 10^n contains no zeros. Therefore, we can find a 10^{-n} such that the product $10^{-n} \times 10^n = 1$ mod all the moduli.

Example:

	mod 3	mod 7	mod 11
10	= 1	3	10
10^2	= 1	2	1
10^3	= 1	6	10
10^4	= 1	4	1
10^6	= 1	1	1

	mod 3	mod 7	mod 11
10^{-1}	= 1	5	10
10^{-2}	= 1	4	1
10^{-3}	= 1	6	10
10^{-4}	= 1	2	1
10^{-6}	= 1	1	1

or 1,000,000 = 1 (mod 3 x 7 x 11) = (mod 1001)

Now, if we accept the classical definition of $10^n \times 10^{-n} = 1$, we can form a fractional number, such as 0.05, by the following methods:

Using Moduli

			3	7	11	
The number	10^2	=	1	2	1	Now, given a number
such as	X	=	1	4	1	and multiplying,
we obtain	1	=	1	1	1	the product of X and 10^2.

Thus $X=10^{-2}$

If	5	=	2	5	5	Then multiplying
by	10^{-2}	=	1	4	1	we generate
	0.05	=	2	6	5	As a check, we
multiply by	0.05	=	2	6	5	and obtain
	0.0025	=	1	1	3	and now
multiplying	10^{-4}	=	1	4	1	we generate the
by						
product	P	=	1	4	3	

If we write the number 25 = 1 4 3 we can observe the correctness of the correspondence by comparison.

Division

As noted previously, the arithmetic operation $Q = \frac{y}{x}$, where x and y are residual numbers less than M, is not defined unless Q is an integer. In order to accomplish this division, a theorem followed by a method of implementation, is given which using defined arithmetic operations, solves for $\frac{1}{x}$ yielding a decimal fraction, which, when multiplied by an integer (y) and rounded, yields Q as the resulting integer.

Division Theorem

Let $x \neq 0$, and y be two integers less than or equal to a fixed constant M, and let it be desired to form the quotient, Q, such that:

$$(1)\quad Q = \frac{y}{x}$$

Then it is asserted that

$$(2)\quad Q = y\ \frac{(1)}{M} \sum_{n=0}^{\infty} (r)^n$$

where r, the additive inverse of x, is defined by the following relationships:

$$(3)\quad x + r = M$$

It follows directly from equation (3) that:

$$(4)\quad x = M - r$$

$$(5)\quad \nu \frac{r}{M} \geq 0$$

Next, consider the following geometric series and its sum:

$$x \sum_{n=0}^{\infty} z^n = (x)\ \frac{1}{1-z}$$

This series converges absolutely, uniformly, and unconditionally to:

$$(x)\ \frac{1}{1-z}$$

Where z is bounded by:

$$1 > z \geq 0$$

Since r satisfies the criteria for (x) and $\frac{r}{M}$ $\frac{(1)}{M}$ is a fixed constant:

$$Q = y\ \frac{(1)}{M} \sum_{n=0}^{\infty} \frac{(r)^n}{M}$$

As a special case of the above, if $M = 10^N$ then:

$$Q = \frac{y\ 10^{-N}}{1 - \frac{r}{10^N}} = Y\ 10^{-N} \sum_{n=0}^{\infty} (r\ 10^{-N})^n$$

where r is the 10 N's complement of x. If binary fractions are used, r is equal to the two's complement of x.

One method for implementing the division criteria is to normalize x as follows:

> Convert x to mixed radix code and simultaneously define x mod the adjoined moduli. Sense leading mixed radix 0's, multiply x by the moduli in which they are defined and store this multiplier to correct M; i.e., if k equals the product of the modulus multipliers of x then:
>
> $$Q = kM \quad = \frac{yk}{kx} \quad \text{where } M = \frac{y}{kx}$$
>
> The next step is to subtract kx from 10^N to define r and using a wired matrix, evaluate in one pulse time:
>
> $Z_o = (1 + 10^{-N}r + 10^{-2N}r^2)\ 10^{2N}$ and use Z_o as the first approximation for the Newton-Raphson iteration which states:
>
> $$\lim_{i \to \infty} Z_i = \frac{1}{x}$$
>
> where Z_{i+1} is defined $Z_{i+1} = Z_i\ (2 - xZ_i)$
>
> which for $r < \frac{M}{2}$ will converge in 3 iterations.
>
> Example of normalization and generation of r:
>
> For x = 40, convert x to mixed radix codes:

	13	11	7
40	1	7	5
-5	8	6	2
35	9	2	0
1			
14	2	8	
5	5	5	
-5	8	6	
0	0	0	
-0	0	0	
0	0	0	

$$40 = 0 \times 77 + 5 \times 7 + 5$$

Since there is a 0 in the leading digit of the mixed radix code Mod 13, multiply 40 in modular notation by 13:

	13	11	7
40	1	7	5
X13	0	2	5
520	0	3	4

To find r, subtract 13 X 40 from 1000:

	13	11	7
1000	12	10	6
-520	0	8	3
r = 480	12	7	9

CONCLUSION

This paper attempts to summarize the residual arithmetic algorithms known to date. Further study in this field will either result in simpler algorithms or determine whether the algorithms presented are, in some sense, the best possible or the simplest. Some investigations are being made in the field of logical design, and preliminary results indicate that all algorithms are logically feasible using conventional hardware. One memory matrix, consisting of $(n/2)^2$ cores, can be used to add, subtract, multiply and perform the operation of $(x - y)^k$ or any function of two variables. Finally it should be noted that the time required for multiplication and division is dependent upon the number of steps that are necessary to establish numbers in a positional coded system. If a faster method than n-1 steps where n is the number of moduli, can be realized, residual arithmetic promises to become the fastest way of doing arithmetic known to date.

BIBLIOGRAPHY

1. H. L. Garner, "The Residue Number System", paper given at Western Joint Computer Conference, June 1959.

2. Uspensky and Heaslet, "Elementary Number Theory", McGraw Hill Book Company, Inc., 1939.

3. A. Svoboda, "Rational Numerical System of Residual Classes", Stroje Na Zpracovani Informaci, Sbornik V; 1957.

A NOTE ON THE APPLICABILITY OF ERROR CORRECTING CODES

James E. Palmer

Applied Research
Defense Electronic Products
RADIO CORPORATION OF AMERICA

INTRODUCTION

In recent years there has been a widespread interest in the subject of redundant error correcting codes. From the vast literature available on the subject, one could be led to believe that the use of error correcting codes will solve the major problem facing digital communications systems--namely, the problem of reducing the number of errors occurring during transmission of data.

While it is true that, by performing the proper logical manipulations called for by the error correcting code, errors can be corrected, there is a price to be paid for the error correcting capabilitity. This price is the necessary reduction in information flow caused by the insertion of the redundancy, which, in itself, carries no information.

In order to properly evaluate the performance of an error correcting code, it is then necessary to consider both the increased immunity to errors and the decreased ability to transmit information. Such an evaluation technique will be described in this paper, using several numerical examples of the use of this evaluation technique.

THE METHOD OF EVALUATION

When data is to be transmitted to a remote location, five factors are usually of prime importance in determining the probability of errors:

1. The Signal Power available at the receiver
2. The Noise Power present at the receiver
3. The Modulation Technique utilized for transmission
4. The Keying Rate
5. The Information Rate

Reiger[1], Lawton[2], and Filipowsky[3], among others, have analyzed the problem of specifying system performance in the following manner. The probability of error for a given modulation technique is plotted as a function of the energy contrast E/N_0. E is the energy per digit and N_0 is the noise power per unit cycle of bandwidth.

The energy contrast is related to the received signal-to-noise ratio by:

$$(1) \quad \frac{E}{N_0} = BT\frac{S}{N}$$

where: S = average received signal power
N = average received noise power
T = the digit interval
B = the channel bandwidth

E/N_0 can be also described as the received signal-to-noise ratio when keying at 1/2 the Nyquist Rate.

Equation (1) accounts for all the factors except data rate. When transmitting information with no redundancy, the information rate is equal to the keying rate. However, when redundancy is transmitted with information, the data rate and the keying rate are different. The following discussion proposes a method for comparing coding techniques at a constant information rate.

Let us consider a data communications system which uses a character composed of D information digits transmitted at a keying rate K digit per second. The time to transmit a character is then

$$(2) \quad T = \frac{D}{K}$$

If it were desired to use a redundant error correcting code, R redundant would be added to the D data digits, which would then be keyed at a rate K' digits per second. The time taken to transmit a character now would be:

$$(3) \quad T' = \frac{D+R}{K'}$$

If both systems are to send at the same information rate then:

$$(4) \quad T = T'$$

or

$$(5) \quad \frac{K'}{K} = \frac{D+R}{D}$$

Either the keying rate of the error correction character must be increased, or as more practical, the keying rate of the unprotected character can be reduced.

If the keying rate of the unprotected character is reduced by an amount indicated by equation (5), then the energy per digit in the unprotected character is increased by ΔE where:

$$(6) \quad \Delta E = 10 \log_{10}\left(\frac{D+R}{D}\right) db$$

Thus, the only way to compare the performance of the two transmission methods at equal information rates is to compare the performance of the error detecting code system at E/N_0, with the performance of the unprotected system at $E/N_0 + \Delta E$.

Note: If, instead of slowing down the unprotected character to the information rate of the protected character, one had speeded up the keying rate of the protected character to match the information rate of the unprotected character, the comparison would have been between the unprotected character at E/N_0 and protected character at $E/N_0 - \Delta E$.

Before proceeding to some illustrations of this comparison technique, it will be worthwhile to define some properties of redundant error correcting codes.

REDUNDANT ERROR CORRECTING CODES

For this paper, a code will be said to be E-- error correcting if all errors of order less than or equal to E are corrected. For example, a triple error correcting code implies the correction of all single, double and triple errors.

In theory, at least, such codes can be set up on the following basis. If there are R redundant digits, and each digit has M levels, then there are M^R possible states of the redundant information. This redundant information can at best be identified with M^R possible states of the received message, one of which must be the state where no errors have occurred. Thus, the relation which must be satisfied for a code of length D + R, M level digits to be E error correcting is:

$$(7) \quad M^R \geq \sum_{\alpha=0}^{E} \binom{D+R}{\alpha} (M-1)^{\alpha}$$

As an example, if M = 2, D = 4, and E = 1,

$$(8) \quad 2^R = 1 + (4+R)$$

The equality holds for R = 3, and the code which is generated is well-known Hamming 4, 3 binary code.

However, integral values of R which satisfy the equality cannot always be found. When the inequality holds, there is some "left over" redundancy. M. S. Corrington[4] has shown that for a digit probability of error less than $\frac{1}{(M-1)(D+R+1)}$ the most reliable code will be obtained if this "left over" redundancy is used to correct as many as possible of the next higher order errors.

The redundancy error correcting codes may not be known or may in fact not exist at all. However, they represent an upper limit on the performance of redundant error correcting codes. Equation (7) allows us to postulate limiting error correcting codes which can only occasionally be reached in practice.

ILLUSTRATIVE EXAMPLE

Consider the following hypothetical data transmission system.

(a) Modulation Technique - Octal non-coherent FSK, using matched filters, all errors equally likely.

(b) Character Length - Five Octal Digits

Reiger[1] has shown that the probability (p) of error in the reception of an octal FSK digit is given by:

$$(9) \quad p = \frac{\exp\left[-\frac{E}{N_0}\right]}{8} \sum_{\lambda=2}^{8} (-1)^{\lambda} \binom{M}{\lambda} \exp\left[-\frac{E}{\lambda N_0}\right]$$

Curve A of Figure 1 shows p as a function of E/N_0.

If the digit errors occur independently, the probability (P) of an error occurring in a sequence of five digits is given by:

$$(10) \quad P = 1 - (1-p)^5$$

which is plotted as curve B of Figure 1.

PERFORMANCE OF A TRIPLE ERROR CORRECTING CODE
(OCTAL FSK MODULATION)

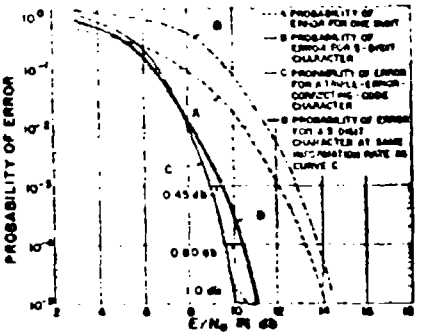


Figure 1

Let us consider the behavior of this system if an "optimum" triple error correcting code as defined by equation (7) is utilized. To determine the number of redundant octal digits which must be added to the original five digit character the following relation must be satisfied:

$$(11) \quad 8^R \geq \sum_{\alpha=0}^{3} \binom{5+R}{\alpha} 7^{\alpha}$$

The smallest value of N which satisfies this equation is z = 6. Using the 6 redundant digits allows correction of all single, double, and triple errors and also allows correction of 205,776 of the 792,330 possible quadruple errors.

The probability of receiving an erroneous character when the error correcting code is utilized is then given by:

$$(12)\quad P' = 1-(1-p)^{11}-11p(1-p)^{10}-55p^2(1-p)^9-$$

$$165p^3(1-p)^8-(330p^4(1-p)^7$$

p' as a function of E/N_0 is plotted as curve C of Figure 1.

A comparison of curves B and C in Figure 1 indicates an apparently large increase in performance obtained through the use of the error detection code.

However, since the coded character requires 11 digits and the uncoded character requires only 5 digits, one could have slowed down the unprotected character to 5/11 of its former rate and produced an information rate equal to the information rate of the coded character. Slowing down the unprotected character increases the energy per digit by the ratio 11/5 (or 3.5 db) and results in curve D of Figure 1.

Comparing the performance of the error correcting code (curve D) with the performance of the slowed down unprotected character (curve D), indicates that the use of the error correcting code produced at best a small increase in performance and at worst actually degrades performance. It is true that the improvement due to the error correcting code increases as the probability of error decreases. This gives rise to the anomaly that "the less one needs an improvement in performance, the more greater the increase in performance provided by the error correcting code."

CODE EVALUATION RESULTS

Hamming 4, 3 Binary Code - Figure 2 shows the relative performance of the Hamming 4, 3 code and an unprotected character of four binary digits, using FSK modulation, or differentially coherent PSK modulation. No significant increase in performance was obtained in range of error probabilities considered.

Hamming 11, 4 Binary Code - Figure 3 shows the relative performance of the Hamming 11, 4 and an unprotected character of 11 binary digits using FSK or differentially coherent PSK. In this case, the code provides an increase in performance of up to 1 db in the range of P_e under consideration.

PERFORMANCE OF THE HAMMING 4,3 CODE
(BINARY FSK MODULATION)

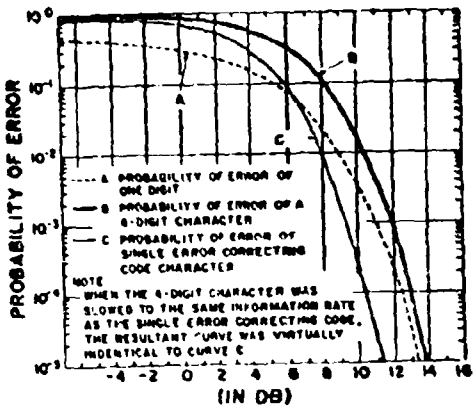


Figure 2

PERFORMANCE OF THE HAMMING 11,4 CODE
(BINARY FSK MODULATION)

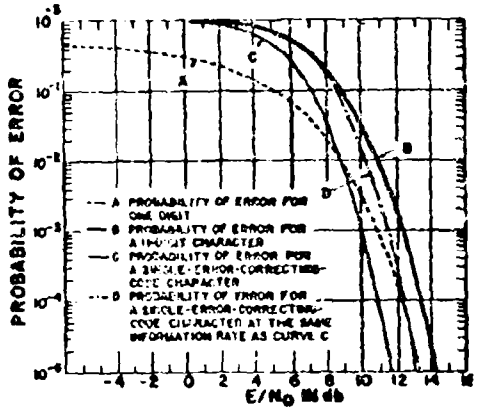


Figure 3

General Binary Codes - The performance of a five binary digit teletype character transmitter by FSK was examined as redundant digits were successively added is shown in Figure 4. The five digit character was somewhat awkward since in the range of values of redundancy considered, none of the codes were close packed and there was always some "left over" redundancy. It is interesting to note that none of the codes considered represent an improvement in performance over the unprotected character.

RELATIVE PERFORMANCE OF REDUNDANT TRANSMISSION TECHNIQUES USING BINARY FSK MODULATION AT A FIXED DATA RATE

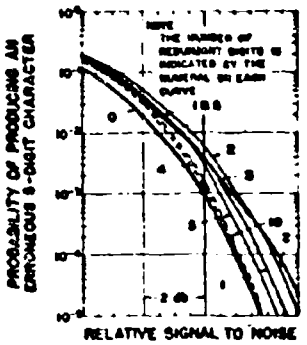


Figure 4

EFFECT OF HIGH REDUNDANCY CODING ON A 5-DIGIT BINARY FSK MESSAGE

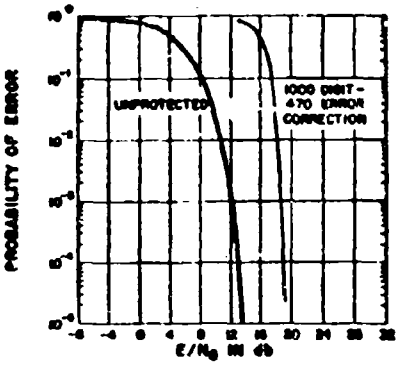


Figure 5

High Redundancy Binary Code - Figure 5 shows the relative performance of a five binary digit character with 995 redundant digits and a five digit unprotected character for FSK. In the range of error probability considered, the high redundancy code represents a severe degradation in performance. It appears that the unprotected character will outperform the protected code at all values of P_e which can be obtained in practice.

Octal Codes - Using five digit octal FSK characters, the performance of single through sexternary error correcting codes was investigated. Figure 7 shows the performance of the various coding schemes at the same information rate as the unprotected character, all the curves lay within the shaded area of the graph.

PERFORMANCE OF ERROR CORRECTING CODES OF VARIOUS ORDERS OF ERROR CORRECTION (OCTAL FSK MODULATION)

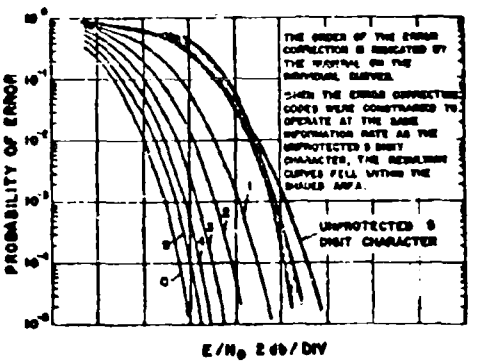


Figure 6

DISCUSSION

Based on the preceding examples and the evaluations of the redundant codes, the following facts were observed.

1. Some redundant codes never gave an increase in performance for any value of P_e.
2. For the other redundant codes, there existed a crossover point at some value of P_e. Above the crossover point, the unprotected character outperformed the redundant code; below the crossover point, the redundant codes outperformed the unprotected character. The increase in performance becomes larger as the probability of character error is reduced.
3. In all the examples investigated, it was questionable whether the small increase in performance justified the increased complexity of the equipment.

REFERENCES

1. Reiger, S., "Some Aspects of Digital Data Transmission," Conference Proceedings, National Conference on Aeronautical Electronics, Dayton, Ohio, 1957, pp. 128-132.
2. Lawton, J. G., "Comparison of Binary Data Transmission Systems," Conference Proceeding, National Convention on Military Electronics, Washington, D.C., 1958, pp. 54-61.
3. Filipowsky, R., "Integrated Data Systems," IRE Transactions on Communications Systems, Vol. CS-7, No. 2 (June, 1959) pp. 95-101.
4. Corrington, M. S., Unpublished Memorandum.

A REAL TIME TELEMETRY DATA TRANSMISSION SYSTEM

By: Mr. H. E. Rennacker, Systems Division, Collins Radio Company, Burbank, California

This paper discusses briefly the need for the real time transmission of telemetry data from remote tracking stations to a central computing facility and then describes a modularly expansible system which can utilize either 3 kc voice lines, or HF radio facilities with a high degree of spectrum conservation.

It also includes remarks on the economic considerations of real time data transmission and the advantages of digital versus analog techniques.

The Need for Real Time Data Transmission

At present, computer input is limited mostly to paper tape and punched cards; far from a real time situation.

The present lack of a rapid data transmission facility imposes a severe handicap on the full utilization of computers which intelligently applied can be much more effective and economical than teams of people pouring over reams of paper.

This handicap will become more apparent after the installation of larger more sophisticated computers.

The transmission of telemetry data over long ground distances in real time for analysis at a central data reduction facility has long been a desirable objective throughout the missile and aircraft industry. However, until the advent of space programs, the cost and complexity of such a system outweighed its advantages. For missile and aircraft flight test evaluation, it was deemed more feasible to record data from the vehicle on magnetic tape and ship the tape back to the home plant for final reduction and analysis. (There, of course, has always been real time recording at the test site of analog data for quick look purposes.)

There have been some notable exceptions such as Convair's use of a leased microwave link for transmission of data from Edwards Air Force Base to the Convair plant in San Diego, and Boeing's use of telephone lines for data transmission between Cape Canaveral and Seattle on the Bomarc program. Such systems were the exception rather than the rule in aircraft and missile testing, however.

Tracking of the first satellites, made the transmission of data in real time over long distances a necessity. The transmission of tracking antenna acquisition data between stations of a tracking network was accomplished by digitizing the output of antenna position servos and by transmitting this data in standard teletype format.

The near real time transmission of payload telemetry data from satellites and space probes to a central computing facility has also become highly desirable if not essential in some cases. This is particularly true for manned vehicles, such as the Mercury Capsule where rapid data analysis and decision making are essential.

Examples of large digital data processing and reduction facilities developed by the aircraft industry are the high speed PCM system developed by CEC and Douglas for use in the DC-8 evaluation and the system which Boeing used on the B-52, KC-135 and 707 programs.

All of these systems, however, operate on data after it has been recorded on magnetic tape; they do not provide for real time data transmission and reduction.

Space craft of the future will undoubtedly, as more powerful vehicles are developed, carry successively more complex and sophisticated payloads.

It is likely also that more tracking and receiving stations will be established around the globe and that more investigative missions will be staged simultaneously.

All of the above add up to an increasing volume of raw telemetry data and a mushrooming requirement for collection, transmission and reduction of this raw data.

In any specific program which is soundly managed, it is assumed that it would be desirable to use information gleaned from earlier missions to revise or redirect the emphasis of later missions. In some cases, expeditious post mortems on aborted missions may be able to identify weaknesses which can be corrected before the launching of later missions and save much valuable time, effort, and money.

As probes are sent deeper into outer space, accuracy requirements for tracking data will become more stringent, in fact tolerances of one cycle of doppler and one thousandth of a degree in bearing have been suggested by cognizant personnel.

Manned space craft will increase the need for real time transmission of telemetry data. Doctors are already commencing to use computers to assist them in improving the speed and reliability of their diagnoses. With the life of a man in space depending on proper and prompt interpretation of telemetered physiological measurements, speed, accuracy and reliability all become more important. Present techniques are inadequate.

Un-manned vehicles attempting to maneuver near or land on other planets or satellites will also require real time transmission and reduction of data.

While general trends are predictable, specific requirements for data transmission and reduction must depend on the programs which are actually assigned and executed.

A Real Time Telemetry Data Transmission System

Before entering into a discussion of the data transmission system described herein, it seems appropriate to review briefly the considerations which led to the selection of this solution.

They may be condensed as follows:

1. Tracking and Telemetry Data should share transmission facilities.
2. The system must be able to handle various types of data.
3. The system must be modularly expandable.
4. Variable accuracy encoding will be required.
5. Upper limit on data transmission bandwidth to be 20 kc.

Finally it was determined that while many in the country have or are expecting to develop problems requiring this type of system, no one has as yet provided the solution.

The system described below meets all of the requirements as reviewed above and exemplifies predicted trends in Data Transmission Systems.

Figure 1 illustrates all of the elements of the basic module of the suggested system.

On the left side of Figure 1 is shown the transmit side of the system which would normally be located at one of the tracking sites.

Starting at the top left of the figure, we find a Frequency Standard and a Synchronizer.

The Frequency Standard in effect is the prime source for the Synchronizer which generates a train of synchronizing pulses which are fed to all of the other units on the left or transmit side of the figure. The function of these pulses will be described later.

The next unit from the top on the left is a Time Converter whose input is time and whose output is digitized time code information to be transmitted periodically with the data.

The third level down consists of from one to eight FM Phase Lock Loop Discriminators, each with its own Frequency Counter and BCD Encoder with a Decimal Display output, a Multiplexer to scan the various outputs, and a programmable BCD to Binary Converter which provides digitized FM telemetry data for transmission.

The fourth item down is a Random Sequence Event Detector and Monitor which monitors selected FM telemetry channels for the occurrence of various anticipated events. On the occurrence of an event the RSEDM which has been emitting zeros, issues an alert code which is followed by the digitized event data.

The last item down is a Tracking Data Statisizer which converts serial tracking data input to a parallel output suitable for transmission.

The next item in the data transmission path is a Pluggable Data Synchronizer and Distributor which consists primarily of a six by eight bit array of core shift registers with downward vertical shifting and right hand horizontal shifting properties. The Gs in the figure represent input and output gates, the Ks represent counters, and the Ps represent parity generators.

Input to the registers is by means of six parallel lines feeding in through the upper or input gates and output is serially through the eight right hand gates and parity generators at 300 bits per line per second. The basic cycling rate of the PDSD is approximately 50 per second.

The eight register lines may be combined in various ways and gates added or deleted as required utilizing the associated plugboard to facilitate revisions. The object is to assign register lines in direct proportion to the bit rates of the inputs.

In the example shown register lines have been assigned as follows with the indicated assumptions. One 300 bps line for synchronization; a synchronizing word being transmitted each cycle. One 300 bps line for time assuming that even though one complete time word would probably be longer than six bits, that time would not be transmitted with each cycle, but no more frequently than once every several cycles.

One 300 bps line has been assigned to handle the output of the RSEDM on the basis that this information rate would be less than 300 bps.

One 300 bps line should be more than adequate for Tracking Data requirements.

Four 300 bps lines were assigned to handle the digitized FM telemetry data on the assumption that the lowest eight IRIG channels are being commutated at the maximum rates and digitized at six bits per sample.

The unit marked TX-300T is a Kineplex transmitter with a capability of handling eight 300 bit per second input channels and using this information to modulate a 3 kc voice channel.

Operation of the system would be as follows:

1. On acquiring the space vehicle as soon as good signals were being received, the operator would introduce a Start signal manually into the Synchronizer which would commence emitting synchronizing pulse trains. The Start signal would reset all counters (K) to zero.
2. As soon as a master synchronizing pulse occurred in each of the four input areas a coincidence gate would open the parallel input gates and the registers would commence to fill, with the synchronizing pulses doing the necessary shifting.

3. Coarse synchronization is accomplished by the proportional assignment of shift register lines on the basis of input data rates. Fine synchronization is provided by the Synchronizer, with blank words inserted as necessary in the input channels to maintain output rates compatible with the transmission equipment.
4. The counters would increment each time a line in the associated register was filled. In this example four of the counters would count to 1 while the fifth would reach a count of 4. The counters are connected modularly as are the register lines, thus the capacity of a given counter would be the same as the number of lines in the associated register.
5. When the counters indicated that their registers were full, the input gates would be closed and the output gates opened to allow the registers' contents to be read out and transmitted by the TE-206T.
6. Parity bits would be generated as the data was read out of the shift registers.
7. As the last information bit left each register line, the output gates would be closed, the input gates would be opened, the counters would be reset to zero, and the registers would commence to fill again.

On the right side of Figure 1 is pictured the receive portion of the data transmission system, normally located at the central facility.

It includes: a TE-206R Kineplex receive unit which demodulates the incoming signal and distributes it over 8 parallel output lines; a Pluggable Data Synchronizer and Statistizer which differs from the transmit PDSD in that its registers have a seven bit width to store the parity bit and they have a serial input, and shift upward and out in parallel.

The Synchronizer and Frequency Standard perform the same functions as their transmit counterparts.

The Multiplexer collects the incoming data and introduces it to the IBM Direct Data Feature which relays it into the IBM 7090 computer.

Operation of gates and counters on the receive side is similar to that on the transmit side.

A new parity bit is generated as the data is read into the receiving registers and compared with the transmitted parity bit which is the last bit received in each line. Parity error indications could be used to signal the computer to disregard the item in error.

In the event that greater than six bit accuracy is required on a particular item, two or more levels or lines may be assigned and the data for the item divided into six bit increments. A seven bit item would be sent in two lines with the least significant six bits sent in one line and the most significant bit sent in the least significant bit position of the next line.

Six bits was chosen as the horizontal dimension of the PDSD on the basis of the most common accuracy requirements, consideration of computer word lengths, timing tolerances, and conservation of space and money.

The system described herein offers flexibility for real time transmission of data ranging from a low capacity single system to a high capacity single system or combinations of systems. This flexibility is achieved merely by adding or subtracting pluggable modules and by changing plugboard configurations.

Economic Considerations of Real Time Data Transmission Methods

Commencing with the requirements that the facilities and methods of data transmission considered must be available, reliable, and modularly expansible to an upper limit of around twenty thousand bits per second, teletype, telephone voice line, and privately owned high frequency radio were selected for comparison.

The three were compared on the basis of overall cost in dollars per month for a distance of one thousand miles versus system capacity in bits per second.

For the fullest utilization of the voice line and H.F. radio facilities, spectrum conserving terminal equipment using orthogonal predicted wave detection techniques was selected. Other systems considered provide less data transmission capacity for equivalent cost.

Costs of all purchased equipment have been amortized over a sixty month period. Teletype and phone line cost estimates were obtained from A. T. and T. Company.

The results of the comparison are shown graphically in Figure 2.

Points X_1, X_2, and X_3 are the capacity thresholds for the respective systems. For the sake of simplicity it is assumed that increases in capacity to the right of these points will be made modularly for each system.

It will be seen from the graph that teletype costs start at $1662 per month for a 75 bit per second system and rise to an exhorbitant $429,500 for a system capable of handling around twenty thousand bits per second.

Kineplex voice line costs start at $6125 for a 2400 bit per second system and rise to only $24,060 for a twenty thousand bit per second system.

Kineplex HF radio costs start at $9485 for a 2400 bit per second link and reach a maximum of $32,390 for the 20,000 bit volume.

Interesting to note is the very narrow capacity range (2.5 to 240 bits per second) in which teletype is superior to a Kineplex voice line

facility.

Taking cognizance of the economies obtainable with spectrum conserving high performance terminal equipment at data rates above 500 bits per second, it appears that it would be better to start with such equipment whenever it is anticipated that data transmission capabilities greater than those of three teletype lines will be required.

Digital Versus Analog Techniques

PM, PAM, and PDM systems have been used in various programs for several years and will probably find continued use because of equipment standardization and availability, and because receiving equipment is available at all test ranges.

Although PCM (digital) systems have been in existence for some years, they have heretofore been impractical because of large size and weight and high cost. (Also the fact that no PCM receiving equipment was available at government test ranges probably delayed use of this system.)

Recent technological developments have enabled the design of new PCM equipments of acceptable size and weight.

The advantages of PCM over analog techniques begin with the fact that PCM exhibits none of the errors due to zero drift, distortion, sensitivity change and other errors which are inherent in an analog system. The criteria for PCM transmission being that signal to noise ratio be sufficient to detect the presence or absence of a pulse. Once this point is reached, the system is independent of noise level.

Variable accuracies are readily obtainable with PCM.

Greater communication efficiency with PCM is a distinct advantage when considering the low transmitter powers available in deep space probes. Variable bandwidth may be used on control from earth to maximize this advantage.

PCM can tolerate tape recorders with poorer frequency response than those required for analog recording.

The use of PCM simplifies the data recognition problem by providing data in a form which can be readily ingested by a computer.

It can be shown that it is more economical to reduce moderate or large volumes of data by digital rather than analog methods.

A more subtle advantage of PCM over analog is that PCM logic lends itself readily to solid state circuitry which provides enhanced reliability.

The 901 Telebit system used in Explorer VI, and the PCM system being developed by Radiation, Inc. for the Minuteman ICBM, are examples of current progress in PCM.

The Digilock orthogonal digital system proposed by Space Electronics Corporation has an even greater communication efficiency than conventional PCM.

Acknowledgment

Acknowledgment is hereby given to R. P. Storey of Collins Radio Company, Alpha Corporation for his many helpful suggestions and useful information drawn from several years of experience in the aviation and missile telemetry fields.

Bibliography

1. Goldman, S., "Information Theory". Prentice-Hall

2. Grief, H.D. and DeKimpe, R.A., "Automatic Data Reduction of Missile Telemetering", 1955 National Telemetering Conference Record, p. 204.

3. IBM Reference Manual, 709 Data Processing System

4. Levy I.J. and Poland, W.B., Jr., "Vanguard Vehicle Telemetry Data-Reduction System".

5. McCormick, E.M., "Data Reduction System for Missile Telemetering", Electronics, Vol. 28, (May, 1955), p. 126.

6. Nichols, M.H. and Rauch, L.L. "Radio Telemetry" Second Edition, John Wiley and Sons, Inc., New York.

7. Tepper, M. "Fundamentals of Radio Telemetry", Rider.

8. "The Doppler Automatic Reduction Equipment", RCA

9. Wynn, J.B., Jr. and Ackerman, S.L., "Guided Missile Test Center, Telemetering System", Electronics, Vol. 25, (May 1952), pp. 106-11

10. "1959 National Symposium on Space Electronics and Telemetry", IRE Proceedings.

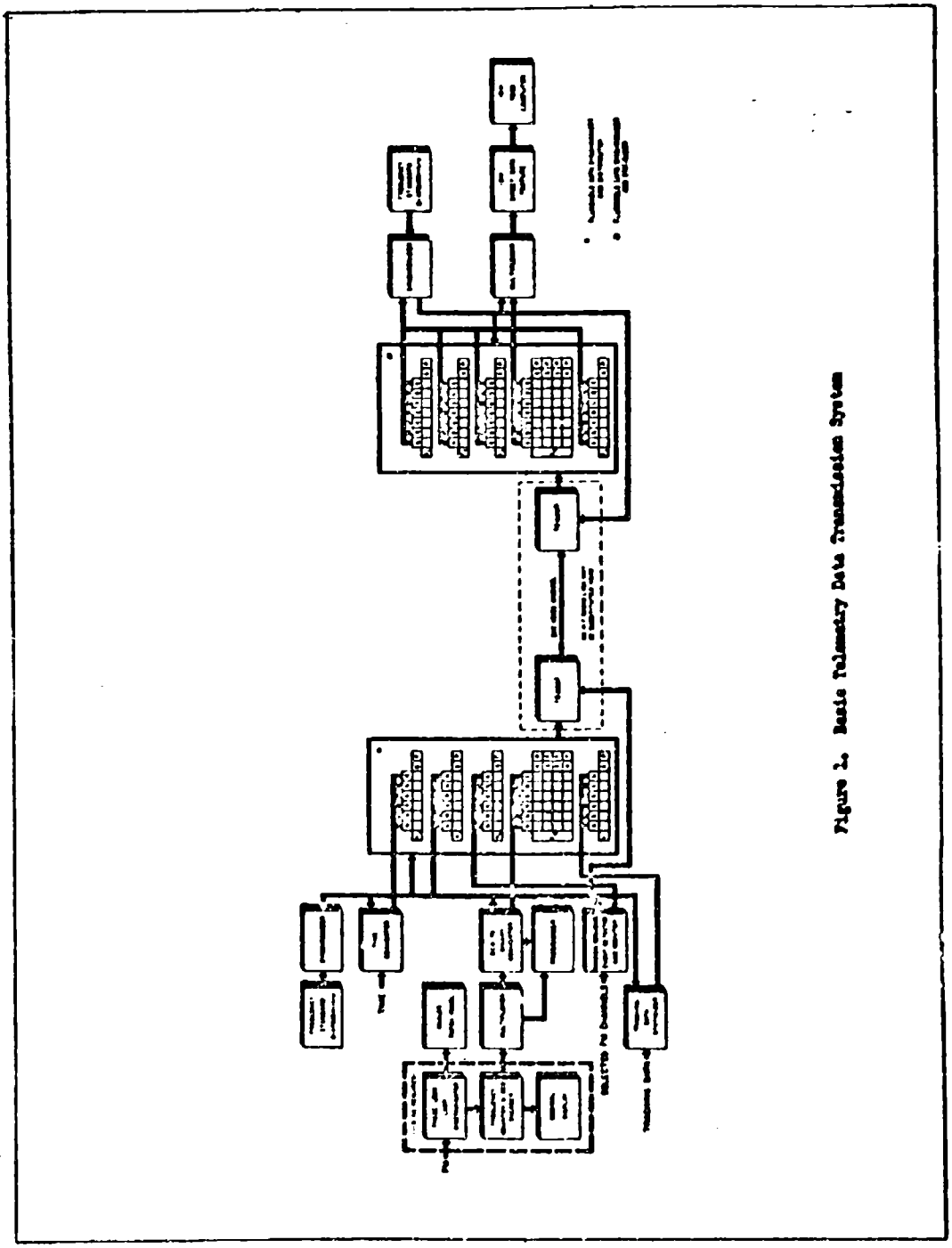


Figure 2. Basic Telemetry Data Transmission System

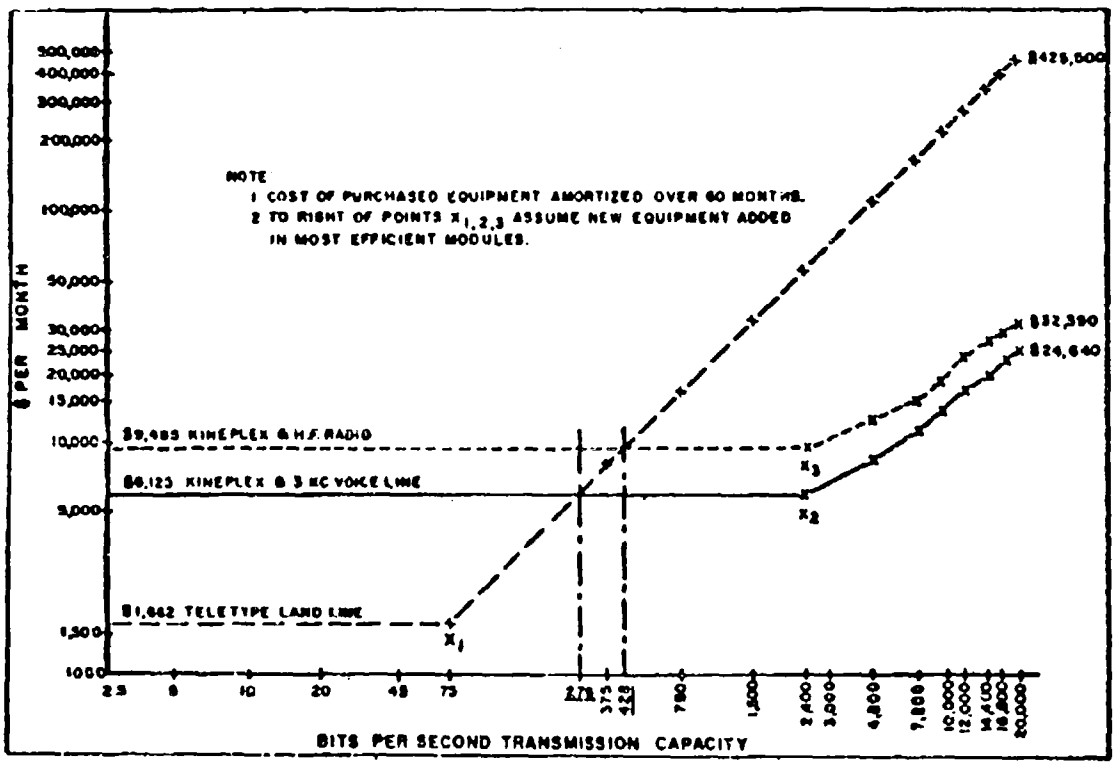


Figure 2. Cost Comparison of Teletype and Kineplex Over a 1000 Mile Run

COMPUTER CONTROLLED AUTOMATIC DIAGNOSTIC AND CHECKOUT SYSTEM FOR FIELD USE

By: Mr. Raymond J. Brachman, Supervisory Fire Control Design Engineer, Frankford Arsenal

Weapons systems in development during the early 1950's are now in the field. Sophisticated weapons systems in development since the mid-1950's are starting to make their way into the field. These contain sub-systems which in themselves are intricate electronic and electro-mechanical assemblies. Despite improvements in reliability, these complex systems still require frequent and a deplorable amount of time consuming maintenance.

While weapons systems show many areas of ingenious design, the maintenance support of these systems has, unfortunately, not received the same attention. Conventional maintenance procedures, utilizing unique and specialized test equipment, operated by high technical skill level personnel, are currently required to maintain these weapons systems. The use of peculiar or unique test equipment for each of these complex systems has become an impossible burden on the military. Maintenance equipment has started to rival the weapons systems in quantity, complexity and cost. The training of highly skilled technicians, the maintenance of a large and complex logistics supply system and the high cost of repair in the field further tend to reduce the overall effectiveness of these weapons systems.

The military effectiveness of these weapons systems is dependent upon rapid and proper maintenance in the field.

Weapons systems today represent the most heterogeneous collection of components, assemblies, assembly techniques, voltage and signal levels imaginable. The Ordnance Corps, with its responsibility to the Army for designing, developing and maintaining this apparently endless variety of weapons systems, had to find a solution to the "impossible" maintenance task. Completely new and advanced thinking in the field of maintenance had to be undertaken in an attempt to keep these highly sophisticated weapons systems operational and tactically useful.

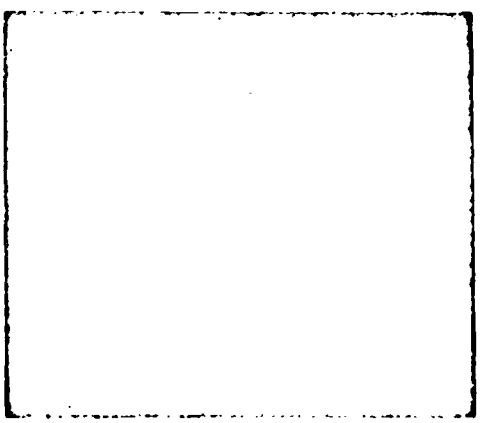
Frankford Arsenal, with background and experience in the field of digital computers and digital control devices, was requested to assist the Army Ordnance Missile Command, its two Agencies, Army Rocket and Guided Missile Agency and Army Ballistic Missile Agency, and to assist the Chief of Field Service, Office, Chief of Ordnance in finding a solution to the maintenance problem.

The basic guide lines were rather straightforward, that is, the test equipment had to be a device that was both universal in application and automatic in operation. Also, the test equipment had to operate with many weapons systems currently in the field and other weapons systems in advanced stages of development without imposing special requirements on the systems to achieve compatibility. Further, the test equipment had to be in itself reliable, reasonably low cost, operable and maintainable by military personnel with low technical skill levels. This device had to be small enough to be highly mobile, survive in the military field environment and provide the test capabilities with sufficient speed to effectively support these weapons systems.

A thorough investigation of all automatic type test equipment then being developed in this country revealed that they were all serial tape or card program systems. These devices still required operators of reasonably high skill level. Fault isolation ability was not sufficient to eliminate the need for highly skilled technicians. The tape programmed universal test set in itself was sufficiently complex as to present its own maintenance problems to its user. Further, these devices required many varied and complex adapters to further complicate the operation of the test set. Last, but by no means least, the programming of these devices was extremely complicated and difficult.

To eliminate the highly skilled technician, the test equipment had to have some analytical capability in addition to its measuring and control capability. The general purpose type digital computer immediately became the ideal device to provide the necessary functions of switching, control, measurement, storage and arithmetic analysis. Frankford Arsenal's experience with the Juke Box and FADAC digital computers showed that the other requirements for operation, maintenance, size and cost could be met.

A program, sponsored by Army Rocket and Guided Missile Agency, starting at Frankford Arsenal in January 1959 and ending the latter part of March 1959, had as its goal proving the feasibility of computer control automatic test equipment. The feasibility breadboard is shown in slide #1.



This equipment contains two 10 x 10 cross bar switches. The capacity of the switching system to connect stimuli to the unit under test was 10 single function lines to any of 50 output test unit lines. The 10 function lines could be connected to any 10 of the 50 output lines at the same time.

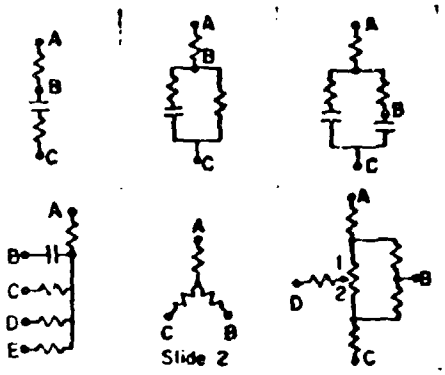
The assigned functions for the 10 lines are

1 - meter +
2 - meter -
3 - voltage +
4 - voltage -
5 - resistance
6 - resistance
7 - short 1
8 - short 1
9 - short 2
10 - short 2

In addition, the measurements performed were limited to resistance and AC and DC voltages. Stimuli was limited to DC, 60 cps, 100 cps and 10,000 cps.

Access to the unit under test was limited to input and output connectors and tube sockets shorting between access points was permitted; however, opening of any circuit was prohibited.

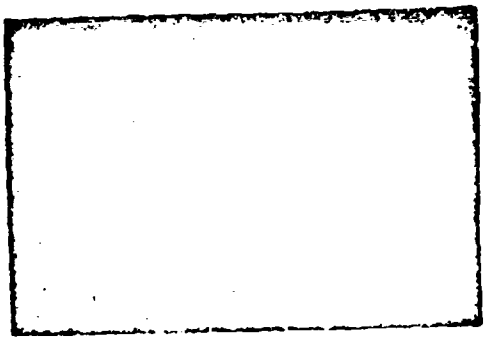
The breadboard equipment was capable of isolating or determining the actual value of all components in the following typical circuits. (Access points are shown by letters and circles). Slide #2



Slide 2

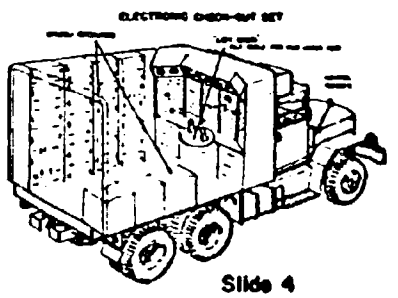
Once the solution of the network was determined, it was treated as a sub-routine. All that was required to solve the network again in a different circuit, was to store the values of the piece parts and their tolerances.

The work was extended to explore more advanced circuit analysis techniques. The breadboard John Box computer was replaced by an LGP-30. The LGP-30 was integrated into the breadboard equipment without modifying the computer. A switching monitor was also added. This is shown in Slide #3



Technical details of this program are available in a Frankford Arsenal Report M 60-5-1.

As a result of this overall program, a computer controller specification has been prepared by Frankford Arsenal and has been included in the Army Rocket and Guided Missile Agency's technical specifications for a field maintenance system under the title, "Multi-purpose Test Equipment." This equipment will provide fault diagnosis to the smallest replaceable component whether it is an assembly, sub-assembly, or piece part. It will be automatic, operable by low skill level personnel, and incorporate to the maximum extent computer control and computer logic for stimuli switching, connector switching and load switching. Measurements have been reduced to numerical conversion of analog data utilizing the computer's internal command structure and arithmetical capabilities to assign voltage, current, resistance, or time values to the data. The computer will also provide the control switching, error detection of proper or improper sequencing, measurement of stimuli, and correction of test values for minor variations in supply readings. The computer, through diagnostic routines, will be capable of checking out the complete "Multi-purpose Test Equipment. Maintenance of the computer proper by personnel with low technical skill levels will be accomplished by the use of a separate device referred to as automatic logic tester. The automatic logic tester is the only piece of equipment in the supporting chain. One logic tester will be able to check itself, to check another logic tester, or the computer. Slide #4 represents an artist's concept of the Multi-purpose Test Equipment. The vehicle shown is an M109 2½ ton shop van. The present system will be packaged in a "Mali-Hut."



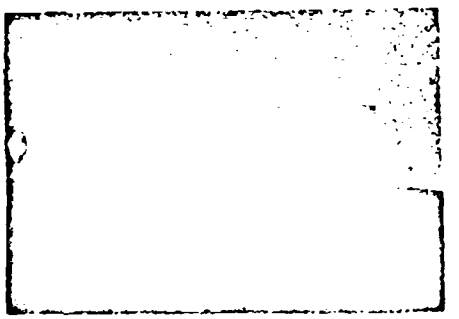
Slide 4

This equipment will be capable of supporting the present missile systems as well as those arriving in the field within the next five years. More advanced systems should require only the changing of the stimuli without any changes to the basic controller.

In an area not related to missile systems, but of equal importance to the Ordnance Corps, is the maintenance of combat vehicles, namely, tanks. A second program is under way at Frankford Arsenal to develop an automatic testing system capable of fault diagnosing combat vehicles. The current program requires the controller equipment or test equipment to provide the analysis of a tank engine, transmission and cooling system. This program is being sponsored by Field Service Division of the Office, Chief of Ordnance. The controller which is currently being fabricated at Frankford Arsenal and which will be supplied to the Letterkenny Ordnance Depot for use in the tank rebuild program is shown in Slide #5

The Ordnance Tank-Automotive Command is currently engaged in a test program to provide the back-up material that will be used as the basis for determining the goodness of an engine, its cooling system and its transmission system. This data will then be reduced and programmed into the computer to provide the means for checking out the vehicle system. Upon completion of the first phase of testing at Letterkenny, the program will be extended to cover the complete vehicle, including items such as power controls, turret controls and fire control.

Future vehicle checkout test equipment will use the controller and switching portion of the missile checkout test equipment. This is in keeping with the Ordnance policy of providing the best maintenance possible without overly complicating the maintenance equipment both from a supply and personnel training standpoint.



SOME ASPECTS OF TUNNEL DIODE APPLICATIONS

By: Mr. Thomas O. Krueger
U. S. Army Signal Research and Development Laboratory
Fort Monmouth, New Jersey

ABSTRACT

Both characterization of the tunnel diode and its application as a single-stage amplifier are discussed. Problems and limitations, especially in military equipment usage, are included as well as performance information. Methods of measuring the various parameters of the diode equivalent circuit are outlined. These methods are primarily adaptations of standard bridge and static techniques, plus measurements from operation as oscillators and amplifiers. The frequency dependence of terminal negative conductance is computed and compared with measurement. Most measurement problems are occasioned by the low diode impedances involved. Formulae predicting operation as small-signal video and band-pass (tuned) amplifiers are developed. These show potentially satisfactory operation to at least 100 Mc with the devices available for experimentation. Noise figures are at present on a par with transistor amplifiers. Burnout characteristics are good; some diodes were subjected to currents as high as one ampere without altering their characteristics.

Many problems exist, however, in a practical application: 1) The temperature dependence of negative conductance may cause unstable operation. Germanium diodes have considerable temperature effect; silicon diodes have less. 2) Small changes in source or load impedance also cause instability. 3) Impedance levels are usually very low (10 to 50 ohms) and create termination and impedance transforming problems, especially if the amplifier is the first stage of a receiver. For the same reason, construction techniques are more difficult. The extremes of operating conditions required of military equipment demand careful consideration in the use of tunnel diode amplifiers.

1. INTRODUCTION

The tunnel, or Esaki, diode is the most recent and perhaps most publicized of all two-terminal negative resistance devices. Several features about it have shown considerable promise both in commercial and military applications. They offer low power consumption, very-high-frequency operation and small physical size. But while a great amount of effort has been expended in developing an understanding of the physical mechanisms and applying the devices to practical circuits, the tunnel diode is still in the early stages of application. It also presents many new and limiting problems to the circuit designer. This paper is intended primarily to acquaint those who are considering using tunnel diodes, particularly in transmission applications, with some of the properties and pitfalls which they will encounter. In military equipment, where stability and consistently high performance are required under widely varying environmental conditions, it is especially important to understand the limitation of the new device.

2. CHARACTERISTICS AND MEASUREMENTS

The tunnel diode is distinguished from a conventional diode by having a negative resistance region of its V-I characteristic (Figure 1).

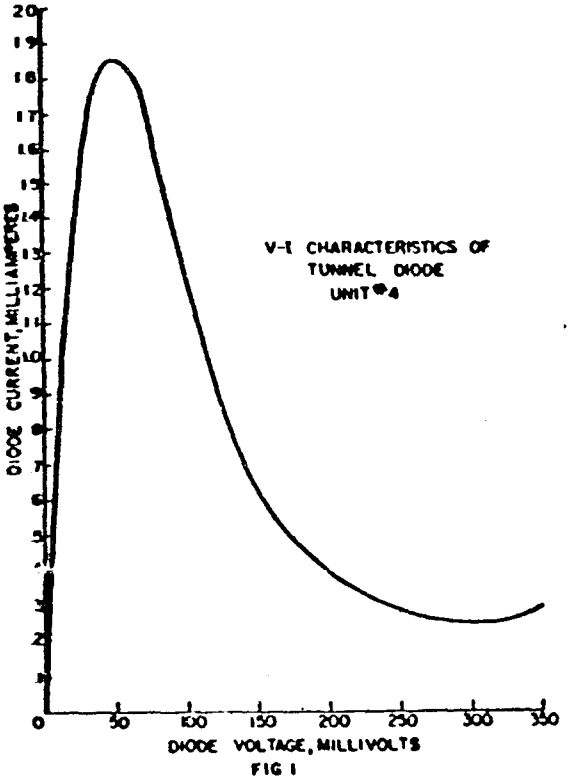
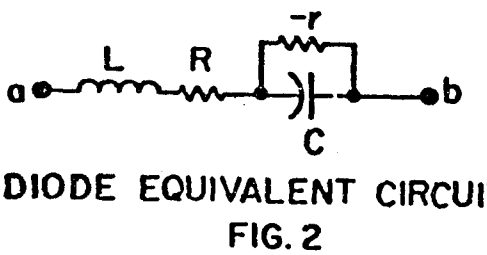


FIG 1

In the negative region, the equivalent circuit of Figure 2 describes the device. The negative resistance is paralleled by the junction capacitance; the bulk semiconductor material contributes the resistance, and the connecting leads, the inductance.



DIODE EQUIVALENT CIRCUIT
FIG. 2

By examining the equation for the series impedance Z_{ab}, we can establish measurement conditions which will produce greatest accuracy.

$$Z_{ab} = \left[R - \frac{r}{1+r^2\omega^2C^2}\right] + j\omega\left[L - \frac{r^2C}{1+r^2\omega^2C^2}\right] \quad (1)$$

For both real and reactive terms, one condition is that $r^2\omega^2C^2$ be very small. From the conversion of the series Z_{eq} to its parallel elements,

$$\left.\begin{aligned} R_p &= \frac{(R+r)^2+(\omega L-r^2\omega C)^2}{R+r} \\ X_p &= \frac{(R+r)^2+(\omega L-r^2\omega C)^2}{\omega L-r^2\omega C} \end{aligned}\right\} \omega^2r^2C^2 \ll 1 \quad (2)$$

a second condition for resistance measurement is that $\omega(L-r^2C)$ should be small compared to $R+r$. In other words, the measurement frequency must be as low as possible.

For capacitance, inaccuracies arise both from L and R. The measured capacitance from equation (2) is:

$$C_m = \frac{L-r^2C}{(R+r)^2+(\omega L-r^2\omega C)^2} \quad (3)$$

This may be separated into factors modifying the actual junction capacitance.

$$C_m = C\left[\frac{r}{r_D}\right]\left[\frac{r}{R+r}\right]\left[1-\frac{LC}{(rC)^2}\right], \quad (\omega^2r^2C^2 \ll 1) \quad (4)$$

The first factor, the ratio of actual to measured negative resistance, is the only frequency-dependent term. The second and third terms include the effects of ohmic resistance and lead inductance, large values of which reduce the measured capacitance. From this it can be seen that most accurate measurements result when r is infinite; that is, when the diode is biased at the peak or valley point.

One source of serious error when trying to make measurements at high frequencies is the series inductance. Very small inductances have considerable effect upon both resistance and reactance. Figure 3 shows the effect upon resistance for five different values of series inductance. Note that even with 5 nH, the curve is much different from the L = 0 case. For this diode, with L = 10 nH and larger, self-resonance occurs. For all L, the cutoff frequency, f_o, remains the same (this is the frequency at which the parallel negative conductance becomes zero). Figure 4 shows the change in capacitance for the same L values. The effect of series inductance on both is such that considerable experimental ingenuity must be applied to reduce circuit and diode package inductance to the point where its effect is minimal. The particular diode described by Figures 3 and 4 is a high-capacitance unit, causing the effects to appear at lower frequencies; the effect is the same for other units.

To measure the diode ohmic resistance, the voltage drop across the junction must either be known or minimized during measurement. It is generally easier to do the latter and sufficiently accurate. The circuit of Figure 5 was used to display the V-I characteristics.

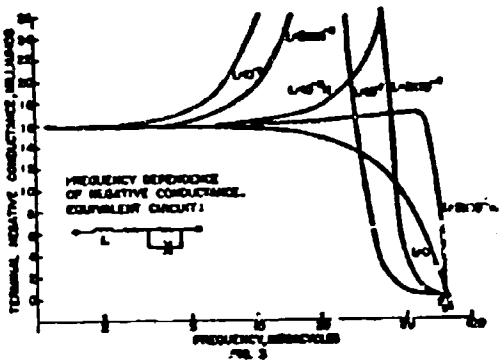


FIG. 3

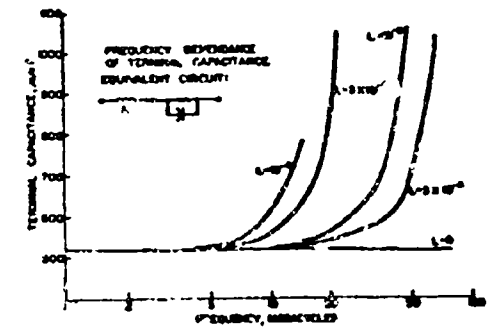
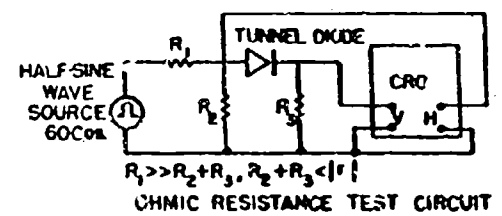


FIG. 4



OHMIC RESISTANCE TEST CIRCUIT

FIG. 5

From the photographed trace, the slope in the forward region at about 50 to 100 times the peak current was scaled to give the resistance. The forward region was chosen because measurements could be made at lower currents and lower power dissipation. In the forward direction the power required to make the junction voltage drop negligible was about half that required for the reverse direction (Figure 6). The fact that for all diodes tested, the same limiting slope was achieved in both forward and reverse directions shows that secondary effects, such as carrier

concentration conductivity modulation, are not important. In making these measurements, it must be remembered that excessive junction heating can alter or destroy the negative resistance characteristic.

Negative resistance may be measured by scaling from the slope of the V-I curve or by amplifier gain or bridge methods. Capacitance may be measured with a bridge or from a circuit oscillation frequency.

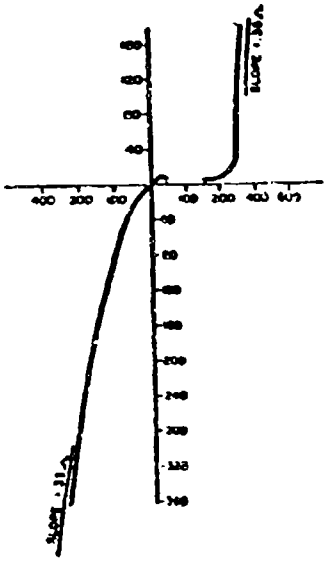
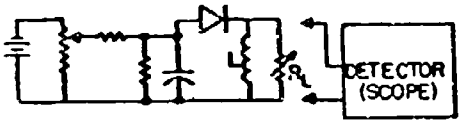


FIG. 6 FORWARD AND REVERSE CHARACTERISTIC FOR MEASURING R SPECIES

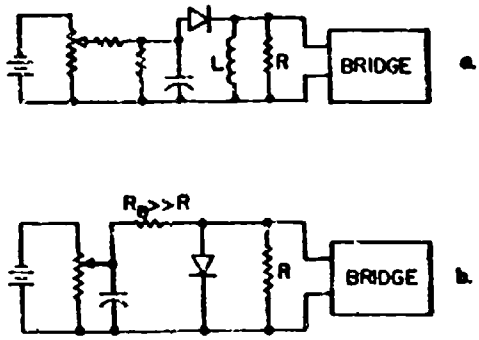
The amplifier circuit of Figure 7 may be used to measure both resistance and capacitance.



AMPLIFIER TEST CIRCUIT
FIG. 7

L is chosen to resonate with the diode capacitance at a low frequency (1 Mc or less). R_L is increased until the circuit is about to oscillate; at this point R_L and -r are of equal magnitude. If R_L is increased further to the point of small-signal oscillations, C may be calculated from the frequency of oscillation. All stray reactances, especially in resistors and capacitors must be accounted for. Accuracy is good; values for both r and C were within five percent of values from bridge measurements. This method can be used only in the negative resistance region, however, and is not easily used at the bias point where -r is minimum. It does offer an alternative method if a bridge is not available.

An RF bridge represents the most satisfactory means for parameter measurement. The Boonton Radio Corporation "RX" meter was found to be very satisfactory in both parameter and frequency range. It is also necessary that the bridge operate at very low signal levels (5 mv RMS or less). The test circuits are diagrammed in Figure 8.



BRIDGE TEST CIRCUITS
FIG. 8

Circuit (a) is for diodes with capacitance beyond the normal range of the bridge, while circuit (b) is for low capacitance diodes. The bridge will then read the difference quantities, R-r and $X_L - X_C$ (or C directly). In a method such as this, accuracy is dependent almost entirely upon the particular physical test circuit. Lead inductance has the greatest effect and must be kept to an absolute minimum.

Device lead inductance is the most difficult quantity to measure, but has a very important effect upon circuit performance as shown. The problem is essentially physical and should be minimized in the diode package. Considering the potentially low junction capacitances, a package with wire leads is not to be recommended. A better approach is to integrate the diode as closely as possible with its external circuit.

Of considerable interest also is the variation of the junction parameters with operating point. Negative resistance is strongly dependent as shown in Figure 9, curve A. At the peak and valley points the resistance becomes infinite, and a minimum occurs generally in the range of 50-100 mv for germanium diodes. The capacitance variation is less straightforward. Entirely different results were obtained from diodes of different manufacture. In all cases a capacitance variation was observed. One example is shown in curve B in Figure 9, over the negative resistance region. Here there is a 20 percent variation over the region, with a minimum occurring at a voltage different from the resistance minimum. Such

results can neither be explained by present published theory or measurement inaccuracies, but they show that cognisance of such a variation by the circuit designer is necessary.

Variations in negative resistance with temperature will be discussed later.

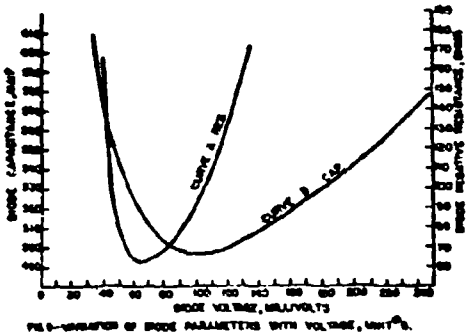


FIG. 5 - VARIATION OF DIODE PARAMETERS WITH VOLTAGE, UNITS A.

3. AMPLIFIER CONSIDERATIONS

Considerable interest has developed in the application of diodes to high frequency amplifiers. But since it is a two-terminal device, the circuit development is considerably different from conventional four-terminal amplifiers. Input and output terminals are identical; unilateral amplification is not possible, per se. This makes amplifier characteristics much more dependent upon the magnitude of source and load admittance. For a given source and load conductance, gain is achieved by an effective increase in voltage across the load with the device, over what the source alone would deliver. Matching conductances to source or load does not necessarily prove advantageous, especially if lowest noise figure is desired. So generally the most significant gains are insertion gains, and if not otherwise noted in the following discussions, insertion gains are intended. Only lumped parameter amplifiers will be discussed here. However, many of the same ideas apply to distributed parameter amplifiers as well.

The important gain formulae for various input and output conditions are summarized in Table A. Because of low impedances and the consequent importance of parasitic elements, it is necessary to use the actual diode and circuit parameters at the operating frequency to accurately predict gain.

For all circuits there are basic frequency limitations.[1] The diode self-resonant frequency, f_r, was discussed earlier. The negative resistance cutoff frequency, f_α, results from the series ohmic resistance of the diode. At this frequency the terminal negative conductance is zero. It may be found from the complete expression for terminal resistance:

$$R_p = \left[\frac{1}{\omega^2 r^2 C_j^2}\right]\left[\omega^2 r + R\omega^2 r^2 C + \frac{[\omega C_j(1+\omega^2 r^2 C_j^2) - \omega^3 r^2 C_j]^2}{R+r+R\omega^2 r^2 C}\right] \quad (5)$$

This will be zero when

$$R + r + R\omega^2 r^2 C = 0$$

at the frequency

$$f_\alpha = \frac{1}{2\pi r C}\sqrt{\frac{R+r}{-R}} \quad . \quad (6)$$

This frequency will usually be higher than the self-resonant frequency f_r. It was pointed out in the discussion of Figures 3 and 4 that f_α is independent of the series inductance.

The amplitude cutoff frequency is associated with the low-pass or video amplifier. This is a low-frequency, generally conductance terminated application. In this case the equivalent circuit may be reduced to a simple parallel of -g and C. For the pure conductance termination, the complex voltage gain is

$$K_v(\omega) = \frac{G_g + G_L}{G_L + G_g - g + j\omega C} \quad . \quad (7)$$

Voltage gain will "cutoff" at the frequency where

$$f_c = \frac{G_g + G_L - g}{2\pi C} \quad . \quad (8)$$

This is a simple R-C parallel circuit and gain attenuation above f_c will be at the rate of 3 db per octave. The gain expression implies a circuit with load and source conductances in parallel, but depending upon diode conductance, it may be preferable to operate the diode, source and load in a series circuit.

Gain and bandwidth of the low-pass amplifier are both expressed in terms of total conductance, so that cutoff frequency may be given as

$$f_c = \frac{G_L + G_g}{2\pi C K_{vo}} \quad . \quad (9)$$

Gain-bandwidth product, then, is

$$GB = K_{vo} f_c = \frac{G_L + G_g}{2\pi C} \quad . \quad (10)$$

Since the limit on the sum $G_L + G_g$ is the device conductance $|g|$, the maximum gain-bandwidth product would be given by

$$GB_{max} = \frac{|g|}{2\pi C} \quad . \quad (11)$$

This represents a useful figure of merit; the diode with the lowest capacitance and highest negative conductance would have the highest gain-bandwidth product. Diodes with very large -g will, however, create additional circuit problems, especially from series inductance.

TABLE A GAIN FORMULAE

Current insertion gain	K_{ic}	$= \frac{G_g+G_L}{G_g+G_L+G-g}$
Voltage insertion gain	K_{iv}	$= \frac{G_g+G_L}{G_g+G_L+G-g}$
Power insertion gain	K_{ip}	$= \frac{(G_g+G_L)^2}{(G_g+G_L+G-g)^2}$
When $G_g = G_L$	K_{ip}	$= \frac{4G_g^2}{(2G_g+G-g)^2}$
Transducer power gain	K_{pt}	$= \frac{4G_gG_L}{(G_g+G_L+G-g)^2}$
Matched output power gain ($G_L = G_g+G-g$)	K_{pom}	$= \frac{G_L}{G_L+G-g}$
Matched input power gain ($G_g = G_L+G-g$)	K_{pim}	$= \frac{G_L}{G_g}$
Matched input and output ($G_g = G_L$)	K_p	$= 1^*$

*This is a degenerate case where $G-g = 0$.

G_g = source conductance

G_L = load conductance

G = loss conductance

$-g$ = diode negative conductance.

Data taken for pure conductance loading show a relatively constant gain-bandwidth product (Table B) for a low-frequency diode.

TABLE B VIDEO AMPLIFIER DATA
(Unit #3)

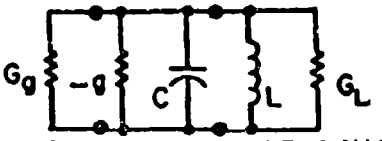
Gain (ratio)	f_c	GB
4.0	1.7 Mc	6.8
6.6	1.0	6.6
9.0	.73	6.6
16	.44	7.0
21	.32	6.7

Other lower capacitance units had gain-bandwidth products of 22 to 58 Mc in this type of circuit. Its use is severely limited by the small output voltages possible and the necessity for non-reactive loading.

The tuned, band-pass amplifier is of considerably more interest. Gain formulae are given in Table A. A short development of bandwidth[2] and noise figure[3] expressions will complete the general discussion.

A suitable equivalent circuit is the simple parallel one of Figure 10. For circuit Q's of greater than about 10, the bandwidth is given with sufficient accuracy by

$$\Delta f = \frac{f_r}{Q_r} = \frac{\omega_r C}{G_g+G_L-g} \quad . \qquad (12)$$

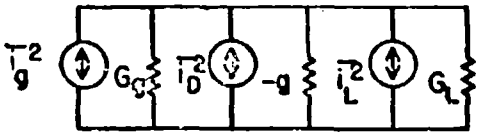


TUNED AMPLIFIER EQUIVALENT CIRCUIT
FIG. 10

However, with high conductance diodes, the circuit Q might be very low. This requires the exact expression

$$\Delta f = \frac{1}{n}(\omega - \omega_r)$$
$$= \frac{G}{C}\left[1+\sqrt{1+\frac{4C}{LG^2}}\right] - 2\omega_r \, . \qquad (13)$$

The noise figure of the one-port amplifier will be defined as the ratio of noise power delivered to an ideal load conductance, G_L, by the noisy amplifier, to the power from the source conductance, G_L, delivered by a noiseless amplifier. The various noise sources are indicated in Figure 11.



ASSUMED NOISE SOURCES
FIG. 11

The noise power delivered to the load by each source is

$$P_{go} = 4kT_s\frac{G_gG_L}{(G_g+G_L-g)^2} = kT_sK_{pim} \qquad (14)$$

$$P_{do} = 2qI_D\frac{G_L}{(G_g+G_L-g)^2} = \frac{qI_D}{2G_g}K_{pim} \qquad (15)$$

$$P_{Lo} = 4kT_s\frac{G_L^2}{(G_g+G_L-g)^2} = kT_s\frac{G_L}{G_g}K_{pim} \qquad (16)$$

(It is interesting to note that the noise power from the input is the product of the matched input noise power and matched input power gain, although no matching condition is implied.) It is assumed that the only noise contribution by the diode is due to shot noise. The noise figure will then be given by the ratio

$$NF = \frac{P_{go} + P_{do} + P_{ro}}{P_{go}} \qquad (17)$$

$$NF = 1 + \frac{G_s}{G_g} + \frac{q I_s}{2kT_s} \frac{1}{G_g} . \qquad (18)$$

The noise figure is therefore not only a function of the diode, but the source and load conductance as well.

In measurements at 10 and 30 Mc, the equation correlated well. No deviation from shot noise characteristics was observed at these frequencies. A representative curve of NF vs operating voltage for constant source and load conductance is shown in Figure 12. Clearly the lowest NF occurs in the direction of higher operating voltage, and lower current. The dashed curve shows the effect of halving source conductance. The dotted curve is for constant source conductance and constant gain. If full advantage of a low-noise diode is to be gained, then the ratio of load to source conductance should be as small as possible. With the germanium diodes available, noise figures as low as 5 db at 30 Mc were measured.

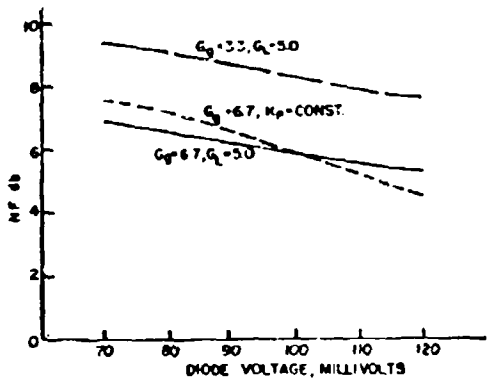


FIG 12 NOISE FIGURE VARIATION WITH CIRCUIT CONSTANTS AND VOLTAGE.

At this point it is necessary to point out some of the problems in achieving satisfactory performance as a tuned amplifier.

Deviations from predicted amplifier performance result from:

1. Stray circuit reactance and resistance.
2. Variation in source and load impedance.
3. Variation in diode parameters.

The effect of series resistance and inductance on effective negative conductance was discussed previously and deserves further emphasis. Most "anomalous" circuit behavior can be traced to this cause. The junction negative conductance is modified by any series R and L; the fact that this conductance is usually high makes normally negligible wiring impedances very important. Parasitic resistance and reactance in any passive component must also be considered. In the extreme case, the circuit will oscillate whenever the real part of the circuit impedance external to the diode junction is equal to or greater than the junction negative resistance. A true negative resistance is not frequency-selective; the external circuit may meet the oscillation criterion at several frequencies and oscillate at these frequencies simultaneously. One circuit was set up which oscillated at three different frequencies. This problem is especially severe when constructing a low-frequency tuned amplifier; parasitic oscillations at a higher frequency are common. In lumped parameter amplifier circuits (and measurement circuits), extreme care must be taken in construction if predictable results are to be achieved.

If reasonable gain is to be expected from a diode amplifier, then the sum of the source and load conductances is very nearly equal to the device negative conductance. Power gain is a second-power function of the difference in these two quantities. Clearly, a small change in any one will produce a large change in gain. For an amplifier with 90-ohm input and output, and a diode -g of 20 mv, the insertion power gain is

$$K_{ip} = \frac{(11.1 + 11.1)^2}{(22.2 - 20)^2} = 20 \text{ db}.$$

If either load or source conductance varies, the result is as shown in Table C. Table D shows the effects of variation in diode conductance. The tables were calculated for only one of three elements, G_g, G_L, and -g variable; of course a practical case might have to contend with variations in all three during operation. They serve to show the magnitude of these effects.

The diode -g can change due either to changes in bias point or temperature. The effect of change in bias can be shown from a plot of conductance and operating voltage (Figure 13). Table E shows approximate change of operating voltage to produce a given percentage change in -g. At 40 mv, a 1.7 percent change in voltage causes a +10 percent change in -g which is nearly sufficient to cause oscillation. At 65 mv, a 6 percent change causes only a 1 percent change in -g. Therefore, depending upon bias point, stabilization of supply voltage may be necessary.

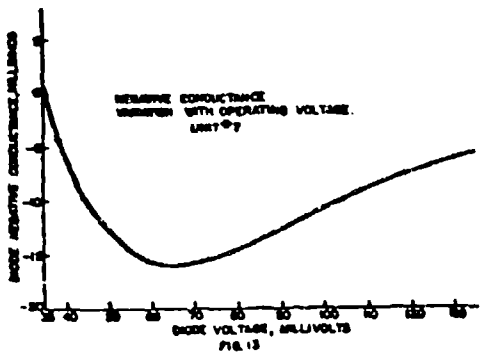


FIG 13

TABLE C GAIN VARIATION WITH SOURCE OR LOAD

G_L or G_g	+20%	-20%	+10%	-10%	+5%	-5%	+2%	-2%
K_{p1}	14.9	*	17.2	25.7	18.4	22.1	19.4	20.5
DIFF	-5.1	-	-2.8	+5.7	-1.6	+2.4	-.6	+.5
%	-26	-	-14	+29	-8	+12	-3	+2.5

*oscillation

TABLE D
GAIN VARIATION WITH DIODE NEGATIVE CONDUCTANCE

-G	+10%	-10%	+5%	-5%	+2%	-2%	+1%	-1%
K_{p1}	40.9	14.5	25.4	16.8	21.8	18.6	20.9	19.3
DIFF	+20.9	-5.5	+5.4	-3.2	+1.8	-1.4	+.9	-.7
%	+105	-28	+27	-16	+9	-7	+4.5	-3.5

TABLE E DIODE NEGATIVE CONDUCTANCE CHANGE WITH VOLTAGE FOR THREE BIAS POINTS

		A -G									
	$V_{op.}$	-20%	+20%	-10%	+10%	-5%	+5%	-2%	+2%	-1%	+1%
	40mv	-1.3	+1.5	-.6	+.7	-	-	-	-	-	-
ΔV	65	-15	-25	-11	+16	-8.5	+11	-5.5	+7	-4	+5.5
	110	-10	+12	- 4	+5.5	-2.5	+3	-1	+1	-	-

Temperature has a very important effect upon the germanium diode, in particular. The variation of -G with temperature for three bias points is shown in Figure 14. Over the range of 27 to 65°C the variation at 90 and 150 mv is approximately constant at .18 percent and .34 percent per degree C, respectively, of the value at 27°C, for the particular diode tested. At 70 mv the -g was roughly constant. This would suggest operating at 70 mv, but noise figure is degraded and sensitivity to bias voltage change is much greater. If germanium diodes are used a compromise must be effected. For large temperature changes, certainly a diode of smaller temperature coefficient such as silicon or gallium arsenide should be used. With these materials and special design methods, it may be possible to greatly reduce temperature variation in negative conductance.

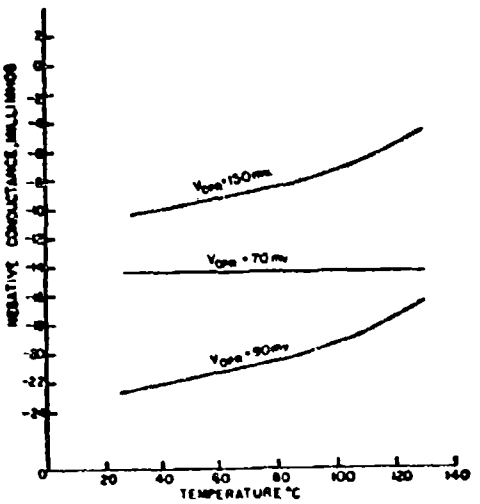
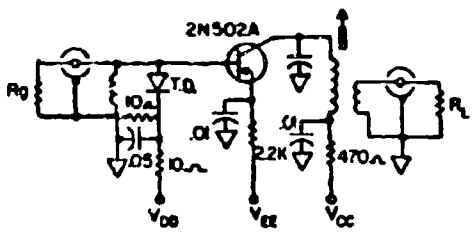


FIG 14-TEMPERATURE DEPENDENCE OF NEGATIVE CONDUCTANCE
UNIT A-14

As one approach to the problem of variations in loading conductances, the diode amplifier was coupled to a transistor isolation stage. First concern was for as little degradation of diode noise figure as possible. For a cascade of this type the total noise figure is

$$NF_T = NF_1 + \frac{NF_A - 1}{K_{p1}} \quad (19)$$

If the transistor is not to materially affect the noise figure, then it too should have a low NF and the gain of the diode should be large. This, however, is not stable. The transistor input conductance (diode load) is also a function of temperature and bias. The conductance should be small compared to the generator conductance to have minimum temperature effect and lowest diode noise figure. The requirement of a low noise transistor is greater because of the disadvantage of the higher source resistance to the transistor's noise figure. In the experimental amplifier (Figure 15), the transistor power gain was 12 db and the diode, 10 db, with a bandwidth of 800 kc at 30 Mc. Overall noise figure was 7 db due primarily to the higher current diode used. Additional compensation of diode gain variation may be achieved by using a temperature-sensitive resistance to vary transistor emitter current and hence input resistance. Variations in source impedance are still a problem, but the isolation from succeeding stages was such that changes of load impedance of ±50 percent had negligible effect on diode gain or noise figure.

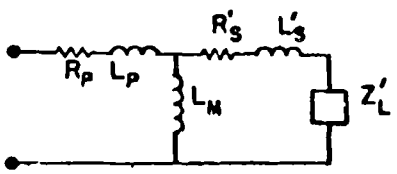


HYBRID AMPLIFIER

FIG. 15

Normally in a tunnel diode amplifier the diode supplies all the capacitive reactance to resonate the circuit at the desired frequency. If external parallel capacitance is used, there is a great possibility of parasitic oscillation. Any physical capacitor has inductance in series with the nominal capacitance and hence a self-resonant frequency. If this occurs below the conductance cutoff frequency of the diode, then the circuit will oscillate. This normally prevents very low frequency tuned amplification, unless very high capacitance diodes are selected.

The diode conductance required in a circuit is determined by load and source conductances, and gain. If the circuit designer does not have a diode of the proper conductance, use of an impedance-changing transformer might be suggested. This, however, is to be discouraged. The transformer will be assumed to have an equivalent circuit as in Figure 16.



TRANSFORMER EQUIVALENT CIRCUIT

FIG. 16

R_p, L_p and R_s', L_s' are the primary and transformed secondary loss resistances and reactances, Z_L is the transformed load impedance. When coupled to the diode, the sources of potential instability become evident. The circuit will either oscillate or show large changes in gain and center frequency as the load impedance is varied. In view of this, use of the transformer is of dubious value except for very carefully controlled loading.

The large-signal action of the diode is similar to any nonlinear active device. Any signal which produces excursions around the operating point in a region of nonlinear -g will, of course, produce distortion in the waveform (Figure 17). If the diode is biased at the point of greatest -g, then the net effect of a large applied signal will be a loss in gain, since any deviation from this point is degenerative (see Figure 13). At other bias points, the excursion causes regeneration in one direction and degeneration in the other, producing a different spectrum of distortion components. When operating near the peak or valley points, the DC component of the distorted waveform may be sufficient to cause the operating point to switch out of the negative region. Additional effects will arise from the nonlinear capacitance variation with operating point, and hence with signal excursion. Particular applications may not tolerate such frequency shifts and distortion products. Very high signal voltages and currents (in one case as high as one ampere) will not damage the diode if the average power dissipation is not excessive.

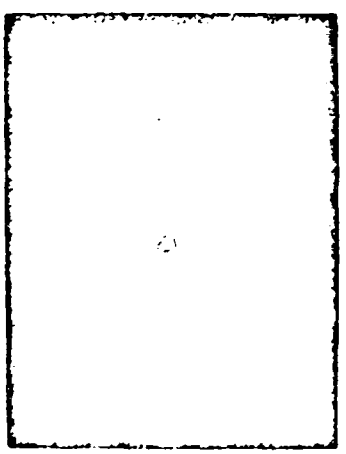


Fig. 17 Large signal distortion
Nominal voltage gain 20 db.
Output voltage top: 220 mv peak-to-peak
middle; 70 mv peak-to-peak
bottom: 10 mv peak-to-peak

There are a number of factors which should be considered in choosing an operating point. First is the value of -g in relation to source, load and gain. Second, lowest noise figures occur at lowest diode current. Third, for stability and large signal operation, the point of least change of -g with source voltage and temperature is desirable. Because of the unavoidable imperfections in transformers, the actual diode -g should be used. If the bias point is chosen for greatest stability, then a diode of proper -g at that point must be selected. (The diode may also be selected for junction capacitance.) This necessitates a low

peak-to-valley ratio diode for low conductance. To minimize noise figure, the current at that point should be as low as possible; the excess current at the valley is the lower limiting factor. Higher resistance diodes are also more desirable because series R and L are less important. This is also the situation of smaller bandwidth. But whatever the AC circuit, the conductance of the DC bias loop must be higher than the desired negative conductance at the bias point, or switching will result. It will generally also be necessary to make the bias voltage adjustable to achieve the desired -g and gain.

The DC power actually consumed by the diode is small, on the order of 50 to 100 microwatts for germanium. In this respect it has an advantage over the transistor which might consume from 1000 microwatts up, to do the same job as the diode. However, millivolt power sources are not readily available. The biasing network required to provide a low impedance, millivolt source to the diode (as from a battery) often consumes ten times as much power as the diode. This presents a challenge to the power source designer to realize the potentiality of a diode circuit in portable equipment.

4. SUMMARY

Three things should be emphasized in making diode parameter measurements. For greatest accuracy, negative resistance and capacitance should be measured at as low a frequency as possible. The AC voltage applied to the diode under test should be very small compared to the operating point voltage. If measurements are to be made at high frequencies, series inductance must be kept to an absolute minimum. Of the three, the last is the most difficult to achieve.

There are many problems in reducing a theoretically practical tunnel diode amplifier to an operating amplifier. The low impedances involved make stray inductance in diode package and circuit construction important or critical. These strays are the greatest handicap to predicting diode amplifier performance. An additional effect resulting from the low impedances is large bandwidth. Smallest bandwidth occurs with high gain, but high gain amplifiers are potentially unstable.

Instability arises from the fact that the diode is a one-port device, where a variation in any one parameter will affect the entire set of amplifier characteristics. Variations in source and load must be analyzed in terms of the particular application. Variations in diode parameters (-g and C) come from changes in operating point or temperature; germanium diodes are most temperature-sensitive.

Noise figure is dependent upon source resistance, load resistance and DC diode current. It does not appear to exhibit a frequency dependence as is observed in transistors.

It is felt that, while lumped parameter high frequency amplifiers are feasible, they do not necessarily represent an improvement over existing transistor amplifiers in the same frequency range.

5. ACKNOWLEDGMENTS

The author wishes to thank Max Schuller for many valuable discussions, especially on temperature effects, and George Silverman for his assistance in preparation of the material.

6. REFERENCES

1. H. S. Sommers, Jr., "Tunnel Diodes as High Frequency Devices," Proc. IRE, 47, p. 1201; 1959.

2. K. K. N. Chang, (Letter) Proc. IRE, 47, p. 1260; 1959.

3. R. Pucel, "The Esaki 'Tunnel' Diode," Technical Memorandum T-171-1, Raytheon Company Research Division; December, 1959.

4. Max Schuller, USAARDL, prive communication.

THE RAYSISTOR, AN ELECTRICAL TRANSFORMER USING OPTICAL COUPLING

By: Mr. James C. Davis, Jr., Industrial Components Division, Raytheon Company

To explain the title: Electro-optical Transformer

This device could be called an amplifier, but its output is not necessarily a sole function of input. It could be called a modulator, or a relay, as it can be used to control a signal or carrier with a control current. I prefer the general term Transformer, as the device transforms electric current from one form to another.

At Raytheon we have for several years been making a multi-filter spectrum analyzer in which the signal spectrum is broken up by resonant filters into some 400 parts and these are sequentially scanned by a commutator.

In very early models the mechanical motion of the resonant elements was "heard" by a high frequency microphone which was scanned over them by a motor. Later an electrical pickoff was built onto each element and these scanned by a rotating switch. Presently, a capacity commutator using a rotating scanning plate and a fixed plate for each element allows rapid scanning without rubbing parts.

None of the above commutation techniques has proven ideal. The first two had short life and were noisy. The last, while better than prior art, still has high insertion loss and requires mechanical motion.

The answer to the problems of this application seemed to require certain conditions be met. First - no moving parts, second - low insertion loss, third - light weight and small size, and lastly, that the analyzer still be as good as the present model.

The result of trying to solve these problems was the "Raysistor".



Basically, what is wanted is a relay capable of very high speeds and using low power. A mechanical relay cannot satisfy these specs because it is too slow and its life, in terms of cycles, is short.

The Raysistor consists of a fast, high-intensity light driving a sensitive, fast, photo-resistive cell. The present unit is not good enough for Rayspan Spectrum Analyzers but is good enough to look interesting for other purposes. Meanwhile we are developing faster, more sensitive models which will do the job.

Let us call a "Raysistor" Relay a Rayswitch. A most obvious advantage of a Rayswitch is the fact that since the drive element is a glow tube, with an associated high resistance, it is very simple to drive from either vacuum tube or high voltage transistor switching circuits. The use of a beam switching tube to drive a "Nixie" numerical indicator is well known. By substituting Rayswitches, a ring counter commutator is achieved.

The information commutated is not restricted with a "Rayswitch" as the controlled element is bilateral. Within the limitation placed by shunt capacity around the "off" condition, the unit will handle either A. C. or D. C. Amplitudes from a few millivolts to several volts are switched with equal loss in Dbs.

In the Rayswitch, due to the fact of complete isolation of the signal circuit from the control circuit, series connection is practical. This opens wide vistas for use in matrix and logic circuits for selection of channels by use of several parallel control signals without requiring, adding or other logic circuits before the switch.

An additional use of Raysistors, of less immediate use to our spectrum analyser but of interest to others, is application as an electrically controlled potentiometer. At present, we do not have a linear potentiometer but what we have should be suitable for feed back circuits, using either high-gain or human feedback.

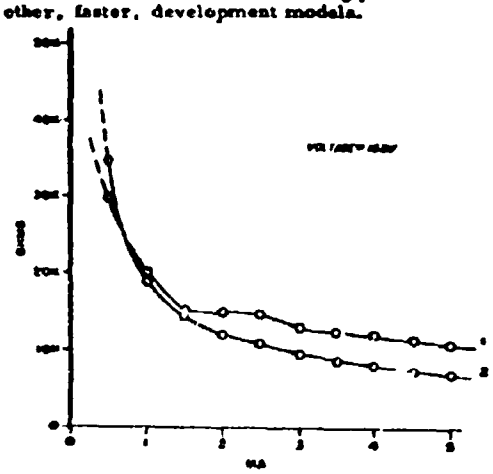
An obvious use is in circuits where a pure resistance is wanted at R. F., for example A. G. C., A. F. C. or Q controls. Unlike a mechanical pot, there is not an appreciable variable capacity associated with the variable resistor.

The shunt capacity of present units can be neutralized if necessary, and future units should have less capacity or lower resistances, therefore reducing capacitive leakage.

Another obvious use is in phase-shifting networks for thyratron or triggered silicon rectifiers. The input-output isolation plus the

fact that D. C. can control A. C. loads to simplification of circuitry.

At present, there is one model of Rayswitch in production. This is what is called a "slow" Rayswitch for it cannot be turned on and off more than 50 to 100 times a second, depending on how off is off. This unit will, however, work with less switching power than other, faster, development models.



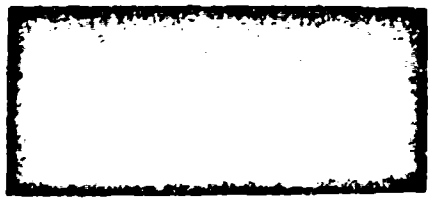
High Impedance Control Raysistor

Speed - 60 cps

Typical values for this Rayswitch are Ic = 3 ma E = 100 v, R = 10 k "on, 10 meg off". Life is not definite but should be better than 10,000 hours.

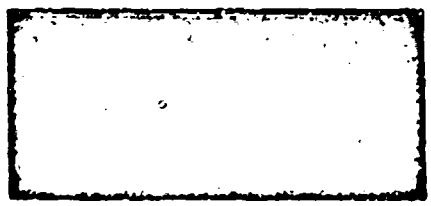
Shunt capacity to R. F. is about 4 pf. If one assumes 60 db attenuation in the "off" condition, this makes the maximum frequency into a 10 k load 5 kc.

We are now developing a "fast" Rayswitch, which will have the following characteristics: speed - on in 2 microseconds, off in 10 to 20, on resistance 1 k , off - better than 1 megohm, capacity, about 1 pf. This unit is far enough along so that these things can be specified. The control power cannot yet be stated.



DRIVE VOLTAGE

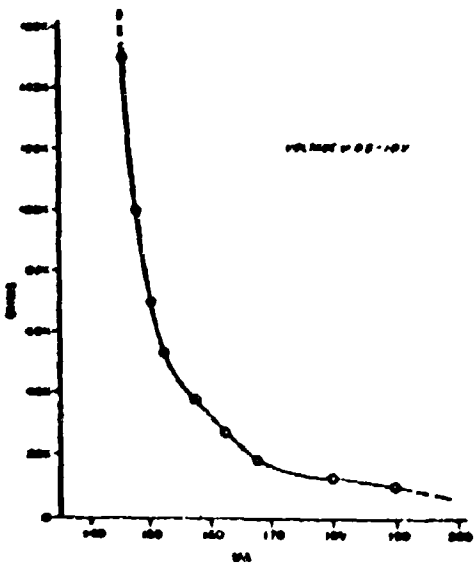
50 μs / square



SIGNAL VOLTAGE

across 10k load

Also under development are linear Raysistors. Present units are suitable for use in feedback circuits but the ultimate hope is a unit at least as good as a triode tube.



Low Impedance Control Raysistor

Speed - 5 cps

THIN FILM COMPONENTS BASED ON TANTALUM

By: Messrs. R. W. Berry & M. Schwartz, Bell Telephone Laboratories, Incorporated

The miniaturization of components and circuitry has become a major development activity of a large segment of the electronics industry. Among the concepts which have evolved is the integrated circuit approach of the Bell Telephone Laboratories. This concept has not dealt with any of the active components, but mainly with resistors and capacitors based on thin films of tantalum and similar metals. Tantalum metal films have a special advantage for component use due to their refractory nature, their ability to form either thermal or anodic oxide films, their inertness to chemical attack and their unusual electrical properties. It is the purpose of this paper to describe the processes involved in making these components, and the properties of the completed capacitors and resistors.

FILM PREPARATION

Although capacitors and resistors have been made from evaporated tantalum films, the majority of the work done to date has been with films prepared by the technique of cathodic sputtering. In brief, this involves creating a low pressure glow discharge in an inert gas with the cathode being made of the material to be deposited. Atomic sized particles are dislodged from the cathode by bombardment of the ionized gas atoms, and deposition of the metal on the substrate is accomplished. Despite the exceptionally high melting point of tantalum, an essentially room temperature deposition of the films can be accomplished by this technique. A wide variety of substrates, including the usual glass and ceramic bases, may be used. The films are extremely adherent, and this is achieved without an extensive pretreatment of the substrates. Part of this adherence is inherent to the mechanism of deposition by sputtering, where an integral cleaning occurs due to the operation of the glow discharge and with the added factor that depositing atoms impinge on the surface with high energies. The resistor films and the capacitor films were prepared under the same conditions, except for the time of deposition. The inert gas used was argon, at a pressure between 10 and 20 microns. The cathode was a disc of N.R.C. arc-melted tantalum four inches in diameter and was placed two inches distant from the substrate which was resting on an aluminum anode. The voltage used during deposition was 4,000V, and the current was 20 milliamperes. Under these conditions, metal was deposited at a rate of approximately 100 Angstrom units per minute. Resistors were sputtered from 5 to 20 minutes, depending upon the resistance value desired, while the capacitor films were usually sputtered for about 45 minutes.

CAPACITORS

The dielectric of the capacitors is anodically grown tantalum oxide and is formed by making the metal films the anode in an electrochemical cell which is completed with an electrolyte (1 pt. oxalic acid, 2 parts water, and 3 parts ethylene glycol, at 105°C has been the main electrolyte for our work) and a tantalum or platinum cathode. Voltage is applied at a current density of approximately 1 milliampere per square centimeter until the desired voltage is reached, and this voltage is maintained on the units for a period of approximately 4 hours. The thickness of the dielectric tantalum oxide formed in this way is directly proportional to the voltage reached in formation, the capacitance then being inversely proportional. In this work, the formation voltage was generally 100 volts although values from 15 to 200 volts have been used.

After the oxide film has been formed, the counter electrode is applied. This is accomplished by vacuum evaporating a thin film of gold or other suitable metal. The structure of the completed capacitor is shown in Figure 1.

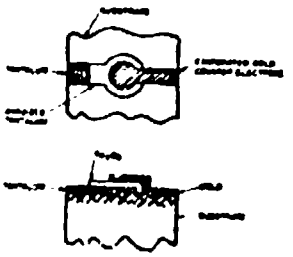


Figure 1

Capacitor Structure

CAPACITOR PROPERTIES

Capacitors produced in this way show outstanding properties which include very low leakage currents, low dissipation factors and high dielectric breakdown

strength. Although the capacitors withstand the full formation voltage without breakdown, maximum operating voltages have not been completely determined as yet. Half the formation voltage, however, has been tentatively selected as a conservative rating for the working voltage. The capacitance-working voltage relationship obtained with these units is best expressed as 5 microfarad-volts per square centimeter. (That is, for a capacitor rated at 50 volts, 0.1 microfarad is achieved using 1 square cm of surface area.) The leakage of these units is extremely small, comparing quite favorably with standard mica capacitors. The room temperature leakage current for a 35,000 picofarad capacitor rated at 50 volts and measured at 75 volts is of the order of 2×10^{-9} amperes. The temperature coefficient is +250 ppm per degree Centigrade, and the dissipation factor is approximately 0.01 at 1 kc and below. The units have shown good stability on life tests with 50 volts applied at 85°C, but it should be pointed out that only limited life test data are available as yet. The properties described here show these units to be quite different from, and quite superior to, normal electrolytic capacitors. Actually, they are electrolytic only in the sense that they have their dielectric film formed electrochemically and should not, therefore, be referred to as electrolytic capacitors.

Although the oxide films usually show the excellent leakage properties described above, occasional units do exhibit high leakage current, and some prove to be short circuits. This behavior is attributed to weak spots in the tantalum oxide film and may be corrected by using a special etching procedure. This is performed in the following manner. The anodized tantalum film (without its counter electrode) is made the anode in an electrochemical etchant for tantalum which does not attack a coherent film of Ta_2O_5 (such as lithium chloride in methyl alcohol). Wherever current can flow, tantalum will be etched away. By applying some voltage slightly below that of formation, the tantalum situated under a weak spot in the tantalum oxide will be removed, and thus the weak spot will be electrically isolated from the capacitor. The unit is then reanodized in order to cover any tantalum exposed in etching. Many units with poor characteristics have been recovered using this technique. Figure 2 illustrates the etching procedure diagrammatically.

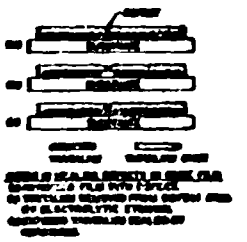


Figure 2

Healing Defects in Tantalum Oxide

The precision of the capacitors is nominally maintained by area and thickness control, although further adjustment may be made by mechanical means. This is possible by designing tabs on one electrode which may be mechanically cut, thereby reducing the effective area. One such electrode design is shown in Figure 3(a). By using the type pattern shown, a total adjustment of -20% is available, with a precision of ±0.5%.

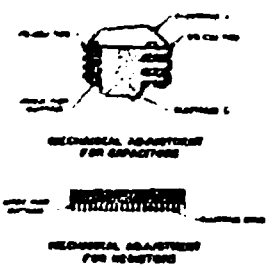


Figure 3

Adjustment Patterns
(a) Capacitors (b) Resistors

RESISTORS

For the basic evaluation of the properties of tantalum resistors, five parallel stripes are sputtered through a mask onto a slide with prefired gold terminations. The dimensions of each stripe is 0.750 inch by 0.050 inch (15 squares) with a 0.25 inch square tab

overlapping the gold termination. The 100 Angstrom units per minute rate of deposition allows fairly close control of the thickness of the films. The resistance can be varied by thickness control over the approximate range of 100 to 10 ohms/square with temperature coefficients falling within the range of ±150 ppm/°C. The temperature coefficients, in common with most metal film resistors, are quite low and differ considerably from the bulk metal. Associated with the low temperature coefficient is a resistivity of 10-15 times that of the bulk metal. These large differences from bulk properties are not primarily functions of the thin film geometry, since the high resistivity is essentially independent of thickness in the range from 500 to 2000 Angstrom units. It is probable that both these unusual properties are related to chemical and structural features within the films. Vacuum fusion analyses indicate an oxygen content in the films of approximately 0.5%, and low angle electron diffraction studies generally indicate an essentially amorphous film structure. This probably originates from the island mechanism usually encountered in the growth of thin films.[1] No detailed correlation of these features to the observed electrical properties of the films has been made, but this aspect of the problem is currently being pursued. The influence that oxygen has in determining the resistivity and temperature coefficient is more than likely complex and probably involves both impurity scattering and a determining influence on film structure.

As with most thin metal films,[2] a considerable instability of resistance is observed for freshly deposited films, both at room and elevated temperatures. This is illustrated in Table 1. The changes in resistance shown in this table

TABLE 1

Per Cent Change in Resistance of Films at 200 Hours

Temp °C	Nominal Thick., Å	$\Delta R/R_o \times 100$
26	500	3.9
	1000	2.8
	2000	1.1
52	500	3.5
	1000	1.8
	2000	1.0
100	500	5.5
	1000	4.8
	2000	1.7

can be attributed to both an annealing of film structure and the growth of an oxide film. It is possible, however, to improve the stability of resistors by preaging in air at 200°C for three hours. This preaging probably results in the growth of a dense adherent thin oxide film. This oxide film apparently acts as protection against further degradation from the atmosphere. The only coating that may be required, then, is one to prevent gross mechanical damage to the film. This marked improvement caused by the preaging is shown in Table 2, where the change in resistance is less than 0.5% in 1000 hrs. at 100°C and, in general, approaches 0.2 to 0.1%.

TABLE 2

Per Cent Change in Resistance at 100°C After Preaging in Air at 200°C for 3 Hours

Thickness, Å	$\Delta R/R_o \times 100$ 200 Hours	1100 Hours
500	-0.05 to +0.15	-0.1 to +0.19
1000	+0.09	+0.19
2000	-0.5	-0.5

The use of tantalum, with its capability of being anodized to form dielectric oxide films, offers the possibility of providing a precision resistor for microcircuitry. By anodizing, approximately 7 Angstrom units of tantalum can be converted to insulating oxide for each volt applied to the electrochemical cell. This then gives very fine control of the effective thickness of the film. The resistance of the metal film can be monitored during the anodizing process, and any predetermined value of resistance can be achieved. Although in the early stages of development, significant results have already been achieved using this method. In one experiment, stripes were sputtered to approximately 1500 Angstrom units with an initial resistance of 110 ohms. By anodizing to different voltages, various values of resistance were achieved up to 2000 ohms. A test of the precision with which a desired resistance value could be reproduced was made by individually anodizing 8 resistor stripes. The anodization is similar to the method described for the capacitors above except that it was performed at room temperature and at constant current, the resistance of the stripe being monitored manually with a Leeds and Northrup Type 8 Test Set. The anodization was discontinued when the resistance reached the predetermined value, and the resistors were removed from the cell and remeasured on an accurate Wheatstone bridge. The data in Table 3 summarize the

initial resistance values, the resistance after anodizing, and the percentage deviation from the mean value and the predetermined value. The results show that

TABLE 3

Test of Precision of Adjustment to 1000 Ohms by Anodization

Init. Res. ohms	Fin. Res. ohms	% Deviation from Mean	% Deviation from Predet. Value
105.79	1012.9	+0.65	+1.29
101.75	1010.1	+0.35	+1.01
98.13	1003.9	-0.25	+0.39
100.38	1004.5	-0.19	+0.45
106.73	1001.6	-0.48	+0.16
100.52	1001.9	-0.45	+0.19
102.50	1011.0	+0.56	+1.10
110.12	1004.4	-0.20	+0.44
Mean	1006.4		

all the resistors were adjusted to +1.3%, and 50% were adjusted to +0.4% of the predetermined value; also almost all of the resistors were within ±0.5% of the mean final value. With the use of more accurate and in particular an automatic monitoring device, it is not unreasonable to assume that adjustment to ±0.1% is feasible. The use of anodization for adjustment also provides an adherent, abrasion resistant oxide film which is more uniform than the thermal oxide films and which apparently provides improved protection. This feature is demonstrated by the 100°C stability of anodized resistors where preliminary data indicate the stability to be at least four times greater, than resistors which had not been anodized. (Both sets had the same preaging of 2 hrs. at 250°C.) This method provides an easily automated continuous adjustment process with final values and stabilities approaching that of wire wound resistors.

Another method for adjustment of resistance is a discontinuous one which allows a 2 to 1 latitude ratio and a precision of about 1-2%. This method depends on intricate resistor patterns possible with a special masking technique previously described.[3] This procedure consists of laying down a film of copper on the substrate, forming the desired pattern by photo-engraving techniques, thus producing a mask which is integral to the surface. A film of the desired thickness of tantalum is sputtered over the entire surface, and the surface is then etched with the usual chemical reagents used for copper. The tantalum is removed from those areas where it is underlayed with copper but adheres at those portions previously exposed by the photo-engraving method. One of the possible resistor patterns is shown in Figure 3. Feasibility studies have shown that these resistor lines have good definition to 2 mil widths and 2 mil spacings with possibilities of achieving 0.5 mil lines with 0.5 mil spacings. The adjustment is achieved by cutting the shorting loops to increase the path length with a scribe or an abrasive loaded gas stream. These processes again can be monitored to achieve a desired resistance. It should be noted that adjustment by anodization is also applicable to these fine patterns.

The processes described in this paper suggest the capability of providing precision components. These capacitors and resistors can be utilized as individual components and would provide for miniaturization and stability urgently needed for present advanced electronics. In a two dimensional topography they provide all the advantages inherent in microcircuitry. Of more importance, however, is the fact that this microcircuitry can be achieved with elements approaching precision components. For the capacitors, this encompasses a low dissipation factor, exceptionally high insulation resistance, and a precise and stable capacitance; while for the resistors, a low temperature coefficient coupled with stabilities generally associated with wire-wound resistors.

ACKNOWLEDGMENTS

We wish to express our appreciation to the many people who have contributed to the processes and development of these components including H. Basseches, W. J. Pendergast, P. L. McGeough, M. J. Urban, D. Sloan, and D. Farrell. We especially wish to acknowledge the guidance and suggestions of D. A. McLean throughout the course of this work.

REFERENCES

1. G. A. Bassett, J. W. Menter, P. W. Pashly, "Structure and Properties of Thin Films", J. Weby & Sons, New York, 1959.

2. R. B. Belser and W. H. Hicklin, J. Applied Phys. 30, 313-322 (1959).

3. D. A. McLean, WESCON Conference Proceedings, 1959.

THERMOELECTRIC GENERATORS FOR SHORT-DURATION APPLICATIONS

By S. R. Hawkins, Solid State Electronics, Research, Missiles and Space Division, Lockheed Aircraft Corp.

1. INTRODUCTION

As a result of several significant breakthroughs in semiconductor materials during the past fifteen years, the direct conversion of heat to electricity using the Seebeck effect has become practical for power generation, rather than merely as a method for measuring temperature. In most applications, these devices are being designed to operate for long periods of time under steady-state conditions. However, there are certain cases where thermoelectric generators can be used with intense heat sources of short duration such as waste heat from a rocket engine, or intense frictional heat of metal against metal, or metal against air, or perhaps even from an explosive gasoline fire. With this type of heat source a "short-term" generator can be used which requires a small heat sink designed only to hold the cold-junction temperature below a specified value within the operating time.

Thermoelectric generators for use with heat sources of short duration can be relatively simple in design compared with the more conventional steady-state types. Because of the short operating time problems such as diffusion of contact materials at high temperatures, and sublimation and oxidation of thermoelectric materials are not as serious as for long-term, steady-state devices (Ref. 1). Thus the designer has considerably more leeway in his choice of materials, as well as in overall design of the unit. The power/weight ratio of a short-term generator can be made extremely high by proper choice of heat-sink material. Calculations indicate that by use of an aluminum heat sink, 64 watts/lb can be obtained, which is higher than can usually be achieved with chemical batteries. This power/weight ratio is also considerably higher than that which has been achieved by steady-state thermoelectric generators, which are typically in the 1-10 watts/lb range depending upon the application. This paper presents an analysis of thermoelectric generators for short-duration use. For given thermocouple and heat-sink materials, and specified values of operating time and temperature differential, an optimum power output - providing minimum overall weight - can be calculated.

2. ANALYSIS

The theory of thermoelectric power generation and the state-of-the-art in thermoelectric materials has been thoroughly discussed in the literature and no attempt will be made to recount it here (Refs. 2-4). Only those equations which are essential to the present analysis will be repeated.

The essential differences between long-term steady-state and short-term nonsteady-state generators can be seen in Fig. 1. Both types are assumed to use high efficiency n-type and p-type semiconductor materials in the thermocouple legs. In the steady-state type, the heat which passes through the device is dissipated into the cooling medium by fins or some other means. This type of generator is designed so that the heat dissipated is equal to the heat taken from the source minus the electrical power delivered to the load. In the nonsteady-state case, where the heat is applied for

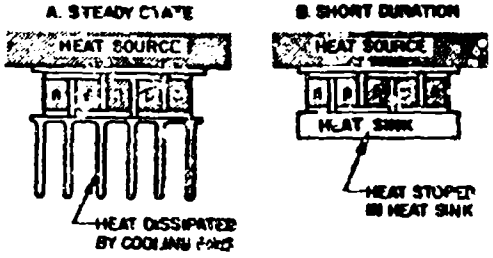


Fig. 1 Thermoelectric Generators

only a short period of time, the heat passing through the generator is stored in a heat sink which slowly rises in temperature as the heat is applied. The longer the heat is applied the greater must be the heat capacity of the sink if it is to hold the temperature below a specified value. If the heat source is very intense, but short in duration, the generator can be designed to produce little or no electrical output when the source temperature rises slowly, by making the heat sink small enough to allow the generator to heat up uniformly so that there will be little or no temperature differential between the junctions. A typical case where this type of design might be useful is one in which the generator is designed to produce a specified electrical output from the intense waste heat of a rocket booster but no output from, say, a warehouse fire.

With a large massive heat sink, the generator could be designed to produce a reverse voltage and current after the intense heat source is removed simply by providing for heat dissipation (e.g., fins) on what is normally the heat collector. In this case, the heat stored in the sink could be made to flow back through the thermocouples to produce a reverse electrical output.

In the analysis that follows, the following assumptions are made:

- The heat source has reached a steady temperature when the electrical output from the generator is utilized.
- The heatup time of the source is small compared with the time of operation.
- The heat sink is designed with a heat capacity large enough to make its temperature variation small compared with the temperature difference between the hot and cold junctions.
- The heat source is large enough that the generator presents no appreciable thermal load.

Under the above assumptions the power generated during source heatup can be considered as a bonus. Any rigorous analysis which takes account of this bonus power would be very difficult due to the complicated temperature dependence of the power output. For this reason, a steady-state type approach is used.

In cases where the heat energy is limited or of extremely short duration, the generator must be designed with a short thermal response time. The time constant of a thermocouple leg can written as $\tau = c_v \rho l^2 / k$.

where

c_v = specific heat
k = thermal conductivity
d = density
ℓ = length

Thus, where a short response time is necessary, the thermocouple legs should be made very short.

2.1 THERMOCOUPLE DESIGN

Figure 2 shows a diagramatic representation of a thermocouple generator as a heat engine. From it we see

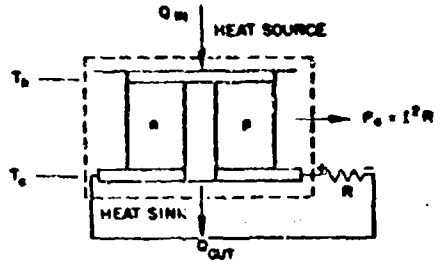


Fig. 2 Thermocouple as a Heat Engine

that the electrical power output is

$$P_e = Q_{in} - Q_{out} \tag{1}$$

where

Q_{in} = heat power taken from the heat source
Q_{out} = heat power deposited into the heat sink

The efficiency may be defined as

$$\eta = \frac{P_e}{Q_{in}} = \frac{\text{power delivered to load}}{\text{power taken from source}} \tag{2}$$

The expression calculated by Ioffe (Ref. 2) for the efficiency of a thermocouple is

$$\eta = \eta_0 \frac{\frac{m}{m+1}}{1 + \frac{1}{Z\bar{T}_h} + \frac{m+1}{2} - \frac{T_h - T_c}{2T_h(m+1)}} \tag{3}$$

where

$\eta_0 = (T_h - T_c)/T_h$ (Carnot Efficiency)
T_h = hot-junction temperature
T_c = cold-junction temperature
$m = R/r$ = ratio of load resistance to generator resistance

The quantity Z is called the device figure of merit and is defined by

$$Z = \frac{\left[n(\bar{\alpha}_n + \bar{\alpha}_p)\right]^2}{r K} \tag{4}$$

where

$\bar{\alpha}_n$ = average Seebeck coefficient for the n-type material
$\bar{\alpha}_p$ = average Seebeck coefficient for the p-type material
r = generator resistance
K = thermal conductance
n = number of thermocouples

Since the Seebeck coefficients are temperature dependent each one should be averaged over the operating temperature range. This is also true for the resistivity used to calculate generator resistance. The temperature dependence of the thermal conductance is usually small and is often neglected.

The maximum efficiency is obtained by differentiating Eq. (3) with respect to m and is

$$\eta_{max} = \eta_0 \frac{M - 1}{M + (T_c/T_h)} \tag{5}$$

where

$$M = (R/r)_{opt} = \left[1 + \frac{1}{2} Z(T_h + T_c)\right]^{1/2} \tag{5a}$$

The matched-load efficiency (where $m = 1$) is

$$\eta_m = \frac{Z(T_h - T_c)}{3T_h + T_c + (8/Z)} \tag{6}$$

Since this is the efficiency at which maximum power is delivered to the load it is more applicable to short-term generators where the heat source is usually large enough that the device presents no appreciable heat load. In cases where the heat source is limited Eq. (5) should be used under the conditions of Eq. (5a) to provide maximum utilization of the heat taken from the source. However the difference between the optimum and matched-load efficiencies is usually small so that a slight mismatch between generator and load may not be serious.

The output heat power of a thermoelectric generator consisting of n thermocouples in series (assuming no losses between the elements) is

$$Q_{out} = \frac{2nkA}{\ell}(T_h - T_c) + n(\alpha_{pc} + \alpha_{nc})T_c + \frac{I^2 r}{2} \tag{7}$$

where

k = thermal conductivity (assumed to be the same for both legs; if not, replace k by $[k_n + k_p]/2$)
A = cross-sectional area of thermocouple leg
ℓ = length of the thermocouple leg
I = current
α_{pc} = Seebeck coefficient of the p-type leg at T_c
α_{nc} = Seebeck coefficient of n-type leg at T_c

The first term on the right of Eq. (7) is the heat loss by conduction while the second is the additional heat which

is liberated into the sink by the Peltier effect. The last term is that part of the internally generated Joule heat which flows into the sink.

Under matched-load conditions

$$I = \frac{2P_e}{E_0}$$

and

$$P_e = I^2 r = I^2 R$$

where E_0 is the open-circuit voltage. E_0 can be expressed as

$$E_0 = n E_{tc} \tag{8}$$

where E_{tc} is the open circuit voltage (Seebeck voltage) from each thermocouple. Substituting these quantities and Eq. (1) and (2) into Eq. (7), we obtain

$$Q_{out} = \frac{P_e}{\eta} - P_e = 2nk(A/\ell)\,\Delta T + (2P_e/E_{tc})(\alpha_{pc} + \alpha_{nc})\,T_h + \frac{P_e}{2}$$

which reduces to

$$P_e = K_0\,A/\ell \tag{9}$$

where

$$K_0 = \frac{2k(T_h - T_c)}{\frac{1}{\eta} - \frac{3}{2} - \frac{2(\alpha_{pc} + \alpha_{nc})\,T_c}{E_{tc}}} \tag{9a}$$

It is often convenient to define the power output per couple from Eq. (9) as

$$P_e' = (P_e/n) = K_0\,A/\ell \tag{9b}$$

It is readily seen that K_0 is dependent only upon the material properties and the junction temperatures. Thus, we see from Eq. (9) that for a given thermoelectric material and specified junction temperatures, the power output from the generator can be adjusted to any desired value which is consistent with the size of the heat source and sink simply by varying the number of thermocouples, or adjusting the A/ℓ ratio of the thermocouple legs, or both. It is usually more desirable from the standpoint of simplicity to adjust the power output by adjusting A/ℓ. However, from Eq. (8) we see that the number of thermocouples is determined by the required output voltage making A/ℓ no longer arbitrary for a given power output. If the required A/ℓ ratio turns out to be impractical, the problem can often be solved by using an appropriate series-parallel array of thermocouples.

The total mass of the thermoelectric material is

$$M_{tc} = 2nA\ell d \tag{10}$$

where d is the density. Combining Eqs. (9) and (10), we obtain

$$M_{tc} = \frac{2n^2 A^2 d K_0}{P_e} = \frac{2K_1}{P_e'} \tag{11}$$

where

$$K_1 = 2A^2 d K_0 \tag{11a}$$

Since the mass or weight of the thermocouples is proportional to A^2, it would seem desirable to make the cross-sectional area as small as possible. There are, however, important limitations as to how small A should be. These can be summarized as follows:

- Since the contact resistance at the junction is inversely proportional to the contact area, A must be large enough to make the contact resistance negligible in comparison to the resistance of the thermocouple leg. Therefore, A cannot be made arbitrarily small to reduce the thermocouple weight without seriously affecting the efficiency of the generator.
- A must be large enough to provide effective heat removal from the cold junctions to avoid creating hot spots in the heat sink.
- Since high efficiency thermoelectric materials are by nature crystalline or polycrystalline and are often brittle, the problem of machining the material to small diameters without excessive breakage can be very serious.

Commercially available thermoelectric materials are usually offered in ingots of various diameters. Therefore, if the material can be obtained with an acceptable diameter, the thermocouple legs can be cut and machined to the desired length directly from the ingot.

2.3 HEAT-SINK DESIGN

The temperature of the heat sink at any given time t be determined from the equation:

$$Q_{out} = M_{hs}\,c_v\,dT/dt \tag{12}$$

where M_{hs} is the mass and c_v the specific heat. By combining Eqs. (1) and (2) we obtain

$$P_e = \eta'\,Q_{out}$$

where

$$\eta' = \frac{\eta}{1 - \eta}$$

Thus we obtain

$$P_e = \frac{M_{hs}\,c_v}{\eta'}\,\frac{dT}{dt} \tag{13}$$

A complete solution to this equation would have to be obtained by numerical methods since both P_e and η' are complicated functions of temperature as evidenced by Eq. (6) and (9). However, if we assumed that the variation of the sink temperature is small compared

with $(T_h - T_0)$, simple integration of Eq. (13) gives

$$P_e(t_1 - t_0) = \frac{M_{hs}\, c_v}{\eta'} (T_1 - T_0) \qquad (14)$$

where

$(t_1 - t_0)$ = operation time of the generator
T_0 = initial temperature
T_1 = final temperature

By combining the above equation with Eq. (13), the mass of the heat sink can be written as

$$M_{hs} = n K_2 P_e' \qquad (15)$$

where

$$K_2 = \frac{t_1 - t_0}{\eta' c_v (T_1 - T_0)} \qquad (15a)$$

The choice of the temperature T_1 depends upon the particular design and how much power the overall device must deliver for the application. However, it is usually chosen so that either the average temperature of the heat sink is equal to the cold-junction temperature T_c used in the thermocouple design, or so that T_1 is equal to T_c. In the latter case the short-term generator will be somewhat overdesigned since the heat-sink temperature will always be less than T_c.

2.3 CHOICE OF HEAT-SINK MATERIAL

Because of the nature of the short-term thermoelectric generator, the choice of the heat-sink material is a very important part of the overall design. In addition to the density, the specific heat and thermal conductivity are very important factors since the heat sink should have a very high heat capacity and no excessive thermal gradients within it. Thermal properties of some possible heat-sink materials are listed in the following table.

Material	Specific Heat $(\frac{w\text{-sec}}{gm\,^\circ C})$	Density x c_v $(\frac{w\text{-sec}}{cm^3\,^\circ C})$	Thermal Conductivity $(\frac{w}{cm\,^\circ C})$
Copper	0.38	3.45	3.80
Aluminum	0.92	2.93	2.09
Magnesium	1.45	1.01	1.57
Beryllium	1.76	3.16	1.80

Copper is the best as far as volume and heat conductivity are concerned but is rather poor with respect to weight. Beryllium would be a very good choice as far as both weight and volume is concerned, and has relatively good thermal conductivity. If the generator is within the specified weight limits, copper is the obvious choice. However, if weight is an important factor other materials should be considered.

A technique for improving heat-sink characteristics recently discussed by Kolenko, (Ref. 5) involves the use of one of the many materials with a large latent heat of fusion having a melting point slightly above room temperature. All or part of the heat is absorbed by the latent heat to fusion without any temperature rise. The time during which there is no temperature increase is $t = Lm'/Q_{out}$ where L is latent heat of fusion and m' is mass of material which becomes molten.

Two of the many possible materials which could be used for this purpose are Woods alloy and sodium phosphate ($Na_2HPO_4 \cdot 12H_2O$). The latent heat properties of these materials are listed below:

	Woods Alloy	Sodium Phosphate
Melting point (°C)	65.5	36.1
Heat of fusion (w-sec/gm)	35.1	279.5
Density x c_v (w-sec/cm^3)	341.0	418.0

Because the heat conductivity of these types of materials is often quite low, the container should have internal fins and be made of a high conductivity metal such as copper to provide even heat distribution.

2.4 OPTIMIZATION OF DESIGN

Since Eqs. (11) and (15) show that M_{tc} decreases while M_{hs} increases with P_e' one would expect that there must be an optimum value of P_e' for which the total mass or weight of the generator is a minimum.

The total mass of the short-term generator can be expressed as

$$M_{tot} = M_{tc} + M_{hs} + nm_0 \qquad (16)$$

where m_0 is the mass per couple of such accessory materials as the contact busbars and the heating plate. It is assumed that m_0 is essentially independent of P_e'. From Eqs. (11) and (14), the total mass per thermocouple is

$$M_{tot}' = K_1/P_e' + K_2 P_e' + m_0 \qquad (16a)$$

To find the power output for minimum weight, differentiate Eq. (16a) with respect to P_e' and set $\partial M_{tot}'/\partial P_e'$ equal to zero:

$$\frac{\partial M_{tot}'}{\partial P_e'} = \frac{-K_1}{P_e'^2} + K_2 = 0$$

The optimum power output per couple is therefore

$$P_{e_{opt}}' = (K_1/K_2)^{1/2} \qquad (17)$$

Substituting $P_{e_{opt}}'$ into Eq. (16a) gives the minimum weight per thermocouple

$$M_{min}' = 2(K_1 K_2)^{1/2} + m_0 \qquad (18)$$

Under these conditions, we must have

$$M_{hs_{opt}} = M_{tc_{opt}}' = (K_1 K_2)^{1/2} \qquad (19)$$

Thus, for an optimized short-term generator, the mass of the heat sink should be

$$M_{hs_{opt}} = a(K_1 K_2)^{1/2} \qquad (19a)$$

From Eq. (16a) we also find the power/weight ratio W to be

$$W = \frac{P'_e}{M'_{tot}} = \frac{P'_e}{\frac{K'}{P'_e} + K_2 P'_e + m_0} \qquad (20)$$

Examination of Eq. (20) shows that W has no finite maxima or minima but approaches an asymptote. Taking the limit of W as P'_e approaches infinity, we find that the asymptote is

$$\lim_{P'_e \to \infty} W = W_{max} = 1/K_2 \qquad (21)$$

which is the maximum power/weight ratio attainable.

From Eqs. (10) and (11), the length of the thermocouple legs for a given value of P_e is found as:

$$\ell = \frac{AK_0}{P'_e} = \frac{K_1}{2AdP'_e} \qquad (22)$$

Substituting Eq. (17) into this, we find that the optimum length of the thermocouple legs is

$$\ell_{opt} = \frac{(K_1 K_2)^{1/2}}{2\lambda d} \qquad (23)$$

We see from Eq. (23) that ℓ_{opt} is independent of both the number of thermocouples and the total power output of the generator. Thus, once the properties of the thermoelectric materials, the junction temperatures, the cross-sectional area of the thermocouple legs, the heat-sink material, and the time of operation are specified, using thermocouple legs of length ℓ_{opt} will optimize the short-term thermoelectric generator for minimum weight, regardless of the number of thermocouples or the total power output.

3. TYPICAL DESIGN CALCULATIONS

To demonstrate the principles outlined in the preceding sections, we shall apply them to design of a short-duration thermoelectric generator to meet a typical requirement. The thermocouple legs will be n-type and p-type lead telluride (PbTe). This material is, at present, the best available commercially for generator applications. It has a maximum operating temperature of about 650°C. The matched load efficiency η_m and the thermocouple voltage E_{tc} for typical PbTe thermocouples are plotted as a function of the hot-junction temperature in Fig. 3 for T_c = 50°C. Since the overall efficiency of typical thermoelectric generators is proving to be about 25% less than that of the materials (due to contact resistance and thermal losses between the thermocouple legs) the efficiencies shown in Fig. 3 have been reduced by this amount to make the following design more realistic. Typical cold-junction Seebeck coefficients for PbTe are as follows:

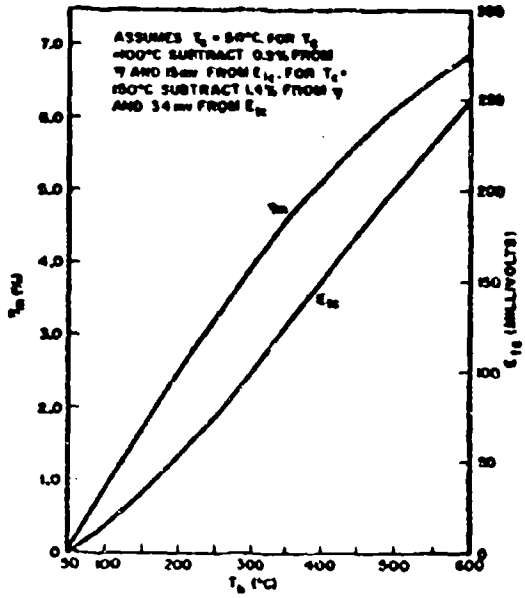


Fig. 3 Efficiency and Thermocouple Voltage

T_c (°C)	α_{nc} (microvolts/°C)	α_{pc} (microvolts/°C)
50	160	84
100	200	107
150	225	134

Applying Eq. (9a) to the data given above and in Fig. 3 gives values for K_0 which are shown in Fig. 4. A thermal conductivity of 0.02 w/cm °C is assumed for the PbTe.

With this information let us design a short-term thermoelectric generator for the following requirements:

- Power output at matched load – 3.5 watts
- Hot-junction temperature – 600°C
- Initial heat-sink temperature – 27°C
- Operating time – 30 sec
- Heat-sink material – copper or aluminum

Let us assume that we are using 1/4-in. diameter n-type and p-type thermocouple legs ($A = 0.31\ cm^2$), and allowing the final heat-sink temperature to be 100°C, which will also be our cold-junction temperature T_c.

From the above conditions and from Fig. 4, we obtain K_0 = 1.35 watts/cm for each couple. Using Eq. (11a) $K_1 = 2A^2 d K_0$ = 2.15 watt-gm/couple. From Fig. 3, η_m = 0.06, making η' = 0.0625. From Eq. (15a):

$$K_2 = \frac{t_1 - t_0}{\eta' c_v (\overline{T_1} - \overline{T_0})} = 17 \text{ gm/watt for copper}$$

$$= 7.1 \text{ gm/watt for aluminum}$$

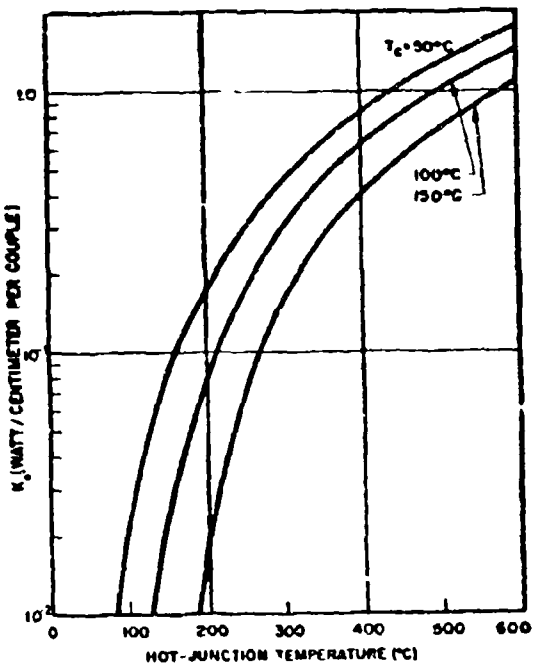


Fig. 4 Watts/cm per Couple for PbTe

where

c_v = 0.39 watt-sec/gm °C for copper
= 0.92 watt-sec/gm °C for aluminum

From the analysis of Section 2 and assuming that m_0 = 10 gm/couple, we obtain the following:

	Copper Heat Sink	Aluminum Heat Sink
$P'_{e_{opt}}$	0.355 watt/couple	0.549 watt/couple
M'_{min}	22 gm/couple	17.8 gm/couple
$M'_{hs_{opt}} = M'_{tc_{opt}}$	6 gm/couple	3.8 gm/couple
W_{opt}	7.5 watt/lb	14.1 watt/lb
ℓ_{opt}	1.19 cm (0.47 in.)	0.77 cm (0.30 in.)
W_{max}	26.7 watt/lb	64 watt/lb
E_{tc}	0.23 volt/couple	0.23 volt/couple

To obtain the required 3.5-watt output from an optimized design we must have ten thermocouples if a copper heat sink is used and six if aluminum is used. Thus, for these two types of short-term generators, we have the following characteristics:

	Copper Heat Sink	Aluminum Heat Sink
n	10 couples	6 couples
M_{min}	220 gm (0.5 lb)	107 gm (0.235 lb)
$M_{hs_{opt}}$	60 gm (0.13 lb)	22.8 gm (0.05 lb)
E_g	2.3 volt	1.38 volt

Figures 5 and 6 show the behavior of M'_{tot} and W

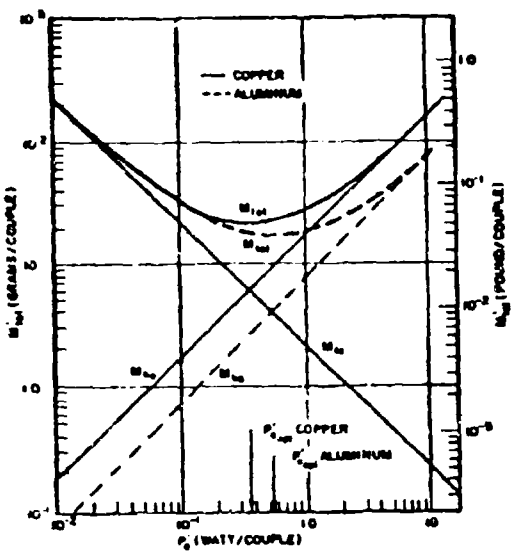


Fig. 5 Weight vs. Power Output

when each is plotted as a function of power output according to Eqs. (16a) and (20) for copper and aluminum heat sinks.

Figure 5 gives an indication of the weight per couple penalty we would pay by not using optimum-length thermocouples. For example, if the number of couples is fixed by a specific voltage requirement, but the power requirement is considerably less than $nP'_{e_{opt}}$ it would not pay to try to design the unit for a power output less than this. In other words, regardless of the required power output, it is always best to use thermocouple legs of length ℓ_{opt}.

Figure 7 is a photograph showing a laboratory model of a short-term generator being demonstrated using an ordinary blowtorch. The generator has twenty PbTe thermocouple legs 1/4-in. in diameter and 1/4-in. long. A dc to dc converter using germanium transistors is mounted to the heat sink of the short-term generator. In this demonstration a bank of capacitors totalling 12,000 μf is charged to about 60 volts in one minute. The weight of the generator alone is about 0.69 lb. A plot of T_h, T_c and P_e for this unit without the converter is shown in Fig. 8 using the blowtorch as an

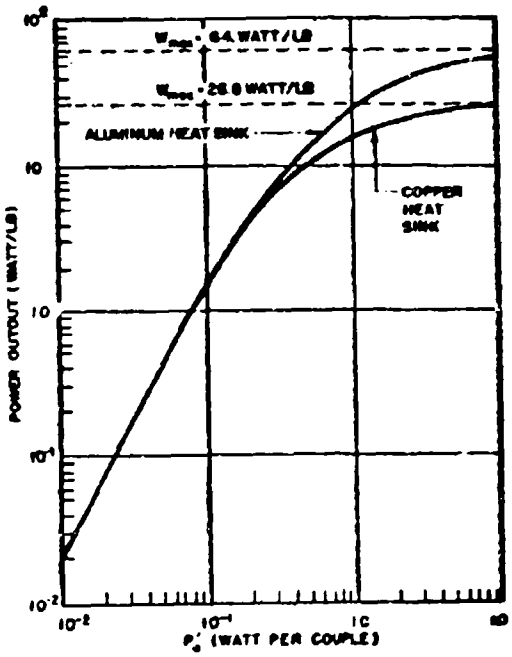


Fig. 6 Power/Weight Ratio vs. Power Output

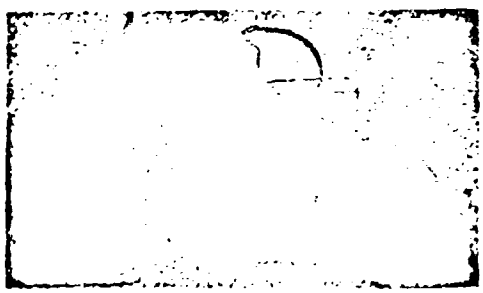


Fig. 7 Demonstration of Laboratory Model

intense heat source. The converter was removed from the heat sink for this test.

ACKNOWLEDGEMENT

The author expresses his appreciation to K. F. Cuff, T. R. Nisbet, C. F. Kooi, and P. S. Castro for their many helpful criticisms and to R. J. Jaffe for his editorial assistance in the preparation of this paper.

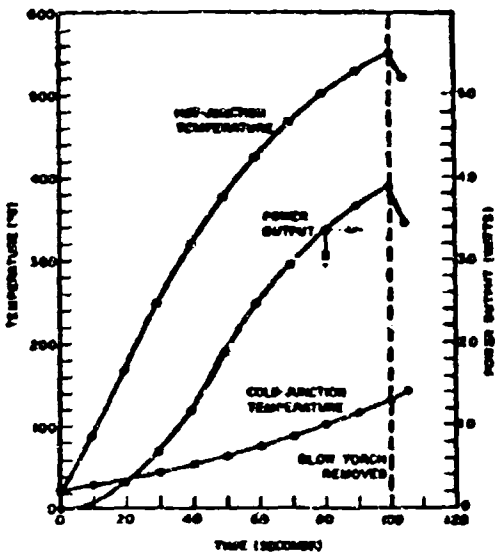


Fig. 8 Performance of Laboratory Model

REFERENCES

1. General Electric Co., Properties Affecting the Utility of Thermoelectric Materials, by P. Klein, Report No. R5 OEL5-11, Electronics Laboratory Technical Information Series, Syracuse, N.Y., Jan. 1960

2. A. F. Ioffe, Semiconductor Thermoelements and Thermoelectric Cooling, London, Infosearch Ltd, 1957

3. B. Sherman et al. "Calculation of the Efficiency of Thermoelectric Devices," J. Appl. Phys., Vol. 31, No. 1, 1960, p. 1

4. R. W. Fritts, Design Parameters for Optimizing the Efficiency of Thermoelectric Generators Utilizing p-type and n-type Lead Telluride, Amer. Inst. of Electrical Engineers Transaction Paper No. 59-909, AIEE, New York, 1959

5. E. A. Kalenko, et al., "Method of Heat Removal from Thermoelectric Cooling Devices, "Zhur. Tekh. Fiz., Vol. 28, No. 11, 1958, pp. 2543-2545

NEW DEVELOPMENTS IN THE FIELD OF MILITARY QUARTZ CRYSTALS

By: Dr. O.E. Gottwein
U.S. Army Signal Research and Development Laboratory, Ft. Monmouth, N.J.

Quartz crystals have been used as frequency controlling elements for more than forty years and have provided reliable operation. Because of this, the opinion is common that all physical and engineering problems connected with them have been solved. This is far from the truth. The continuously stiffening requirements for precision and stability of frequency control elements to be used in modern communications and navigation systems and for space exploration have necessitated new developments in the quartz crystal field and a number of new crystal types are now available which were not in existence a few years ago.

Figure 1 shows a picture of the most commonly used military crystal types. They consist of thin round pieces of quartz plated with metal electrodes which are connected by means of piano wires to the pins in the mount. These crystals are mechanical resonators which are excited to thickness shear vibrations. The thickness is the frequency determining dimension; the thinner the crystal blank the higher its frequency. Resonators of this type are useable for the frequency range of 800 kc to 200 mc. For lower frequency operation, a number of other crystal types are available vibrating in flexural, extensional or face shear modes. Since most of the recent developments are concerned with thickness shear vibrators, the subject of this paper will be restricted to quartz resonators of this type.

The larger crystal can, known as type HC-6, was developed first. The smaller can, known as type HC-18, was introduced later to comply with the miniaturization trend of modern military equipment. The present tendency for newly developed crystal types is to accommodate these resonators, if at all possible, in the size of the HC-18 can. This requires a size reduction of the crystal plate which is easier to accomplish for the thin high frequency crystals than for thicker crystal plates. At present, crystal units for frequencies as low as 3 mc are becoming available in the HC-18 size by means of Signal Corps sponsored Production Engineering contracts.

Crystals of the HC-18 configuration, however, are not the smallest type presently under development. In line with the microminiaturization program of the Signal Corps[1], crystals are being made to fit into the size of a microelement. Figure 2 shows such a crystal unit in disassembled form. The standard wafer size is 0.3" square; the assembled crystal unit is 50 mils thick. The crystal is cemented to the center wafer in 3 points. The two other wafers serve as covers. The maximum blank diameter which this design can accommodate is 0.195 inches. Despite this size limitation crystal units with a frequency as low as 7 mc have been developed for the microminiature configuration.

The crystal blanks are cut from the raw quartz at a certain angle. This angle determines the temperature behavior of the crystal unit, as illustrated in Figure 3. It can be seen that a change in angle of only 10 min of arc has a great effect on the temperature behavior of the crystal unit. Most military crystals maintain their nominal frequency over the temperature range from -55 to +105°C with no greater deviation than ±50 ppm. For a number of applications, however, a better temperature stability is desired. This can be achieved by controlling the angle of cut to a very tight tolerance. By controlling this angle to about 1 minute of arc, it has been found possible to develop crystals maintaining a total tolerance of ± 25 ppm over the temperature range given above. Development of these crystal types has been supported by Production Engineering Measures (PEM) of the Signal Corps and such crystals are now available from a number of manufacturers.

Another area where improvements in stability have been found necessary is to increase the stability versus time. When observed over longer periods of time, all crystals show slight variation in their output frequency, because of a process which is commonly referred to as "aging". The main reason for aging is a change of surface loading of the crystal plate by adsorption or desorption of gases, ejection of small particles, recrystallization processes in the metal plating, and diffusion processes. This can cause either an upward or a downward change in frequency.

Figure 4 illustrates the aging behavior of a group of standard military crystals. All crystals of this group were made by the same manufacturer. They were placed in an oven at 85°C and the output frequency was observed over a period of approximately five months. The curves show that the behavior of the units is quite irregular. In the beginning, some are upward, some are downward. At the end of the observation period, all have decreased in frequency; however, the amount of total aging is quite different for individual crystals.

Two conclusions can be drawn from this curve. (1) All crystals showed an appreciable amount of aging. (2) The long time aging cannot be predicted by the short time behavior of these units over the first ten days.

Future investigation of this matter revealed that the main culprit in the aging behavior is the metal can. It is extremely difficult to clean the cans properly and it is also difficult to seal the cans. Even cans which were tightly sealed during the first ten days showed small leaks after two or three months. Leakage of the cans is the greatest single contributor to a high aging rate.

After appreciable efforts had been expended toward improving the metal cans, this was found to be a difficult task. Although it is probable that by spending considerable time and money a suitable metal can could be developed, it was decided by the Signal Corps that the most promising and possibly cheapest way to obtain low aging crystals was to discard the metal cans completely and to go to another material, namely glass. Glass has always been used as an enclosure material for high precision crystals. Figure 5 shows two glass enclosed crystal units which are equivalent in size and shape to the metal enclosed units of the HC-6 and HC-18 configuration. This makes it possible to interchange glass and metal units. These units are sealed under vacuum. The glass can be easily cleaned and leakers can be detected after the units have been sealed, either by a Tesla coil test or by short time aging tests. Experience has shown that with glass enclosed crystals a good correlation exists between a 5-day short time aging test and a six months long term test[2]. The research and development work on these units is basically completed. Units in HC-6 and HC-18 cans have been developed under Signal Corps contracts, and have met tight aging specifications.

An application where such units will be of vital importance is the field of tactical single side band communications. Experimental pack sets have been built which meet the stringent requirement on frequency stability imposed by suppressed carrier single side band operation. In these sets the crystal units are placed in an adequately controlled oven and operated at 85°C. An aging rate of one part in 10^7 per month appears entirely feasible with these units.

At present, there is no other way to achieve the required frequency stability but to place the units in an oven. This, of course, is an undesirable restriction on size, weight, and power consumption of the equipment since for portable sets, the oven power has to be taken from batteries. Present efforts of the Frequency Control Division of the U.S. Army Signal Research and Development Laboratory are aimed towards minimizing this restriction. Attempts are underway to design an oven with extremely low power consumption. Other research approaches are directed toward finding means for eliminating the oven completely and still maintaining the required frequency stability[3]. All these approaches, however, are in an early development stage and at present not ready for immediate application.

One performance requirement of quartz crystal units which recently has gained importance is their shock and vibration resistance. The present specification calls for a shock test and a two hour vibration test in the range 10 to 55 cycles. For most crystal units, the permissible frequency change after this test is 5 ppm. This test guarantees a certain ruggedness of the crystal design. However, in many applications, particularly for aircraft and guided missiles, exciting frequencies much higher than 50 cps exist. For this reason it is planned to extend the vibration range for newer crystal types to 2000 cps. These requirements, however, are based on "before and after" tests and give no information how well the crystal will maintain its frequency during the vibration excitation. Latest requirements, particularly for missile guidance systems, specify frequency or phase stability during the vibration encountered at the launching of the missile.

To design crystal units meeting these requirements, it is first necessary to examine the mounting structure. It was found that the conventional wire mounts as shown in Figure 1 have mechanical resonances in the range 125 to 500 cycles. When such mounting resonance is excited, a shift of the crystal frequency occurs. It is necessary therefore to eliminate the mounting resonances. Figure 6 shows a picture of a crystal in a very stiff mount whose first mechanical resonance occurs slightly above 2,000 cycles. This mount was used by the Frequency Control Division for special phase stable crystals required by Jet Propulsion Laboratory (JPL) for a missile guidance system. These crystals were found to be extremely rugged and, on the request of JPL, 54 mc crystals of that type were made by the Frequency Control Division for use in the Explorer satellites where they have shown satisfactory performance.

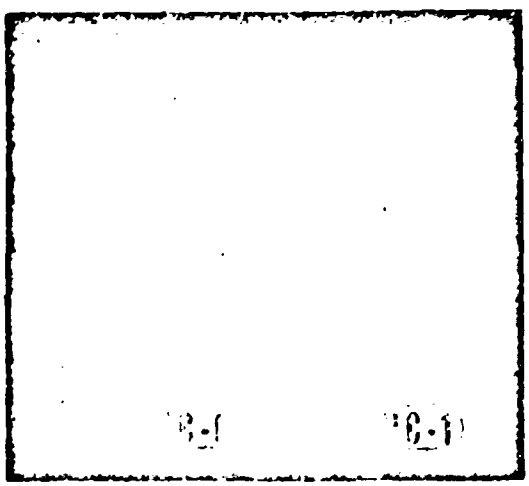
The development of these units, because of lack of time, was mainly done on a cut and try basis. In the meantime, a more systematic investigation has been initiated[4]. One result of that investigation is shown in Figure 7. Plotted here are the measured frequency changes of an oscillating crystal blank under the influence of a compressional force versus the angle of attack of that force. It can be seen that the maximum frequency change for an AT-cut crystal occurs when it is compressed along the X direction. For an angle of 60° off X, no frequency change occurs. When the force is applied along the Z' axis, frequency change is again observed but in the opposite direction. The behavior of another crystal cut, the FT cut, is quite different. It shows smaller frequency changes but it has no neutral axis and the frequency change is always negative.

The existence of a neutral axis as shown in Figure 7 is also of importance for making the crystal output frequency insensitive to static acceleration forces. Investigations conducted by Bell Telephone Laboratories under a Signal Corps contract on the effect of static acceleration on 5 mc high precision crystal units have shown that the mounting of these units is of vital importance[5]. At present, it seems possible with a properly selected mounting to reduce the acceleration coefficient of frequency of these units to about 1 part in 10^{10} per g at least for one direction of the acceleration vector. It is

unquestionable that a considerable amount of research and development work will be necessary before an understanding of the behavior of crystal plates under the influence of external forces is reached and crystals insensitive to vibration and acceleration can be designed to the specifications of the user.

The considerations so far were concerned with the use of crystals as frequency controlling elements. There is another application, however, where quartz crystals are valuable. This is their use as frequency selective elements in filter circuits. Because of their high Q and excellent temperature and time stability, quartz crystals are desirable components for high precision filters. At frequencies below 1 mc, crystal filters have been used for more than thirty years. When it was tried to extend the frequency range of such filters to higher frequencies, particularly to frequencies above 10 mc, great difficulties were encountered because of the spurious mode content of AT-cut crystals. This is illustrated in Figure 8 which shows the spurious mode distribution of a conventional oscillator crystal. The measurement was made by inserting the crystal in a bridge network which was connected to a detector having a logarithmic characteristic. A signal generator was connected to the input of the bridge and was swept through the resonance region of the crystal. The recording shows that a considerable number of spurious modes exist, the strongest of which are only 10 db below the main response. Such spurious modes are intolerable in a filter network. The efforts of the crystal designer therefore were directed toward a suppression of these modes. One of the most efficient means for doing this was developed by the Frequency Control Division by using not circular, but triangular shaped crystal blanks[6]. The improvement obtainable with the triangular shape is shown in Figure 9 which indicates that the strongest spurious mode is 60 db below the main mode. Figure 10 shows a picture of triangular crystals made for the HC-6 and HC-18 holders. The picture also shows a circular filter crystal where the suppression of spurious modes has been accomplished by using a very small electrode size. Although very satisfactory crystal filters can be built by using a crystal design of this type, it is felt that the triangular shaped crystals, because of their better suppression of spurious modes, will become particularly valuable for wide band filters where extreme suppression of spurious modes is necessary to obtain the required pass band.

The above survey shows that, stimulated by the requirements of the user, a very active Research and Development program is being pursued in the quartz crystal field.



References

1. Radio Corporation of America, "Micro-Module Production Program", Contract DA 36-039 SC-75968.

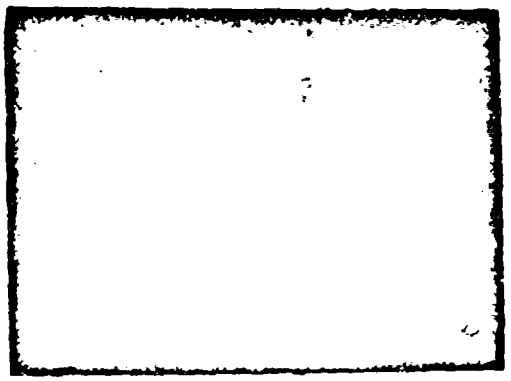
2. P.E. Mulvihill, "Aging Characteristics of Quartz Crystal Units", Proc. 13th Annual Symposium on Frequency Control, 1959, p. 109.

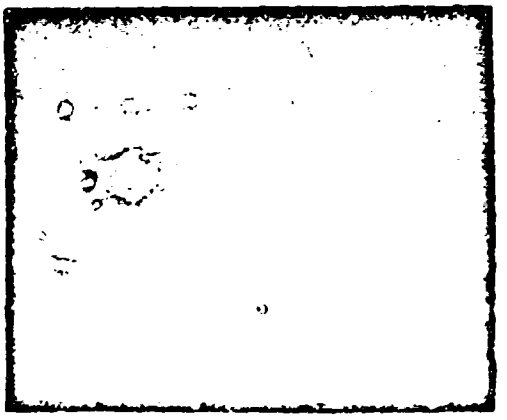
3. E.A. Gerber, "Reduction of Frequency-Temperature Shift of Piezoelectric Crystals by Application of Temperature-Dependent Pressure", Proc. IRE, 48, Feb. 1960, p. 244.

4. A.D. Ballato and R. Bechmann, "Effect of Initial Stress in Vibrating Crystal Plates", Proc. IRE, 48, Feb. 1960, p. 261.

5. A.W. Warner and W.L. Smith, "An Ultra-Precise Standard of Frequency", Contract DA 36-039 SC-73076, 10th, 11th, 12th Interim Report.

6. R. Bechmann, "High Frequency Quartz Filter Crystals", Proc. IRE, 46, March 1958, p. 617.





FREQUENCY TEMPERATURE DEPENDENCE UPON
X — RAY ORIENTATION

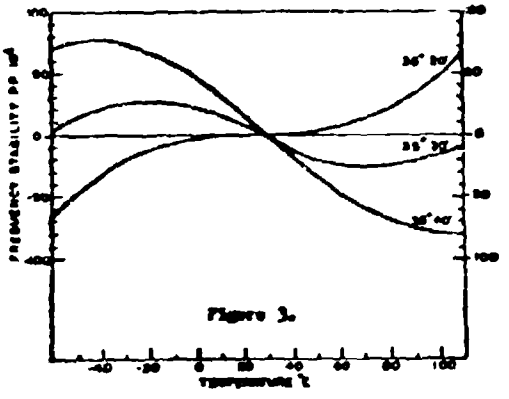
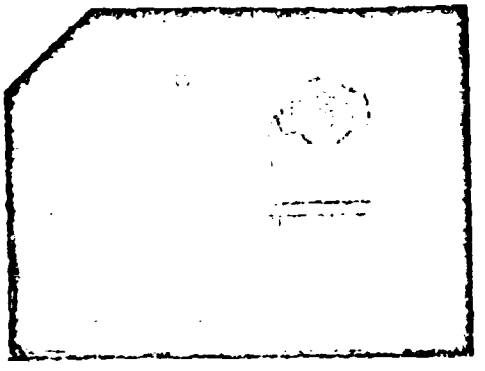


Figure 3.



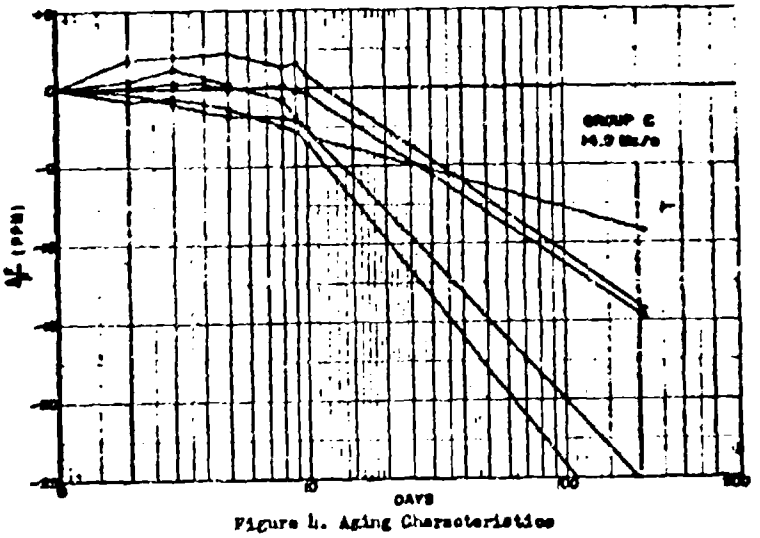


Figure 4. Aging Characteristics

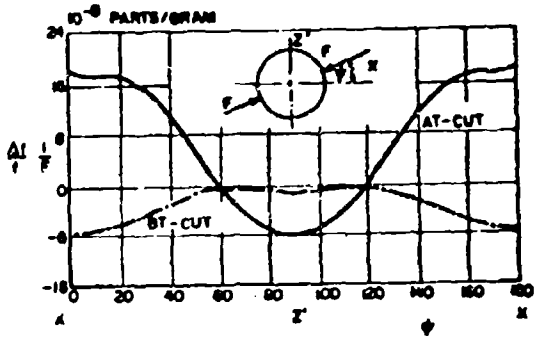


Figure 7

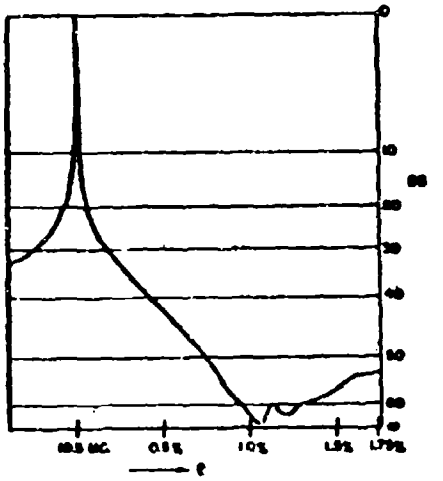


Figure 9 Mode Spectrum of Triangular Filter Crystal

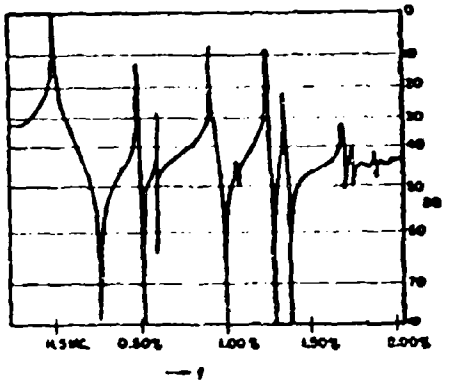


Figure 8

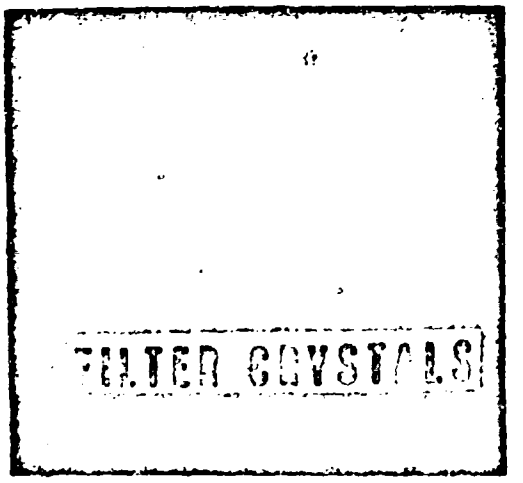


Fig. 10

CERTIFICATION OF COMMAND GUIDANCE MISSILE SYSTEMS

By: Mr. C. M. Redman, White Sands Missile Range, New Mexico

There is considerable concern about standardization of measurements. A volt measured in San Francisco must have the same electro motive force as a volt measured in London. An ohm measured in New York must have the same resistance as an ohm measured in Paris. Even more important, a command guidance missile system tested at Washington must have the same effectiveness as a standard command guidance missile system.

This paper deals with a technique for the standardization or certification of command guidance missile systems. To certify a resistor, a standard of comparison must be available. In the case of the resistor, the standard of comparison may be a resistor certified by the National Bureau of Standards. In the case of a guided missile system, a more complex standard is required. It is difficult to move a guided missile system from location to location to serve as a standard, therefore, it is proposed to record on magnetic tape the performance and use this recorded performance as a standard of comparison.

This would be done as follows: (Figure 1)

(1) Azimuth, elevation and range co-ordinates of the missile radar and target radar would be recorded for typical firing engagements. Commands and certain other pertinent data such as coded time would be recorded with the radar co-ordinates.

(2) The radar co-ordinates would be later played back, compared with the corresponding real-time radar co-ordinates, the differences between recorded co-ordinates and radar co-ordinates operated upon so as to simulate the radar antennas and RF sections and these error signals inserted into the guidance radars.

(3) With the radars tracking the recorded missile and target trajectories, commands and other data from the missile guidance would be compared with the recorded data. At any given time during the engagement, the data should compare within certain tolerances.

A given firing operation can be accurately repeated many times to allow for statistical analysis of the performance. This technique makes possible the following tests:

(1) A command guidance system in any part of the country can be accurately compared with a standard command guidance system in any other part of the country.

(2) Malfunctions can be determined by repeating the exact conditions causing the malfunctions.

(3) By repeating accurately a particular operation, the exact effect of adjustments to a command guidance system can be determined.

(4) A test firing can be repeated any number of times for statistical analysis while expending only one missile.

Melpar Inc., of Watertown, Massachusetts, working on a contract under the control of Frankford Arsenal, has developed a dynamic tester for testing guided missile systems. This dynamic tester is designed to repeat radar azimuth, elevation and range positions with respect to time to an accuracy of 1 part in 64,000, compare these with corresponding shafts on the radar, operate on the above difference so as to simulate the antenna, insert the simulated error into the IF amplifiers and cause the radar to track the program.

The program storage device is an Ampex FR 11b magnetic tape machine. The radar co-ordinates in terms of azimuth, elevation and range are recorded as phase records on the magnetic tape. That is, the zero crossovers of two 400 cycles per second signals are recorded on adjacent channels of the tape to represent each of the co-ordinates. The phase relationship between the recording on two channels represent a co-ordinate of the program with respect to time. The phase relationship of the signals on one pair of channels represents radar range, a second pair represents radar azimuth and a third pair represents radar elevation. These six channels can be reduced to four channels by letting the 400 cycle per second reference signal represent one of each of the pairs. However because of tape errors, the pairs should be adjacent for maximum accuracy.

Figure 1 is a general block diagram of the system. Since each co-ordinate is identical down to the final stages, a description of the elevation angle co-ordinate should suffice. One pair of 400 cycle per second signals representing the elevation angle co-ordinate is played back. One signal (referred to as the data signal) proceeds through amplifiers to a phase detector. The second 400 cycle per second signal (referred to as the reference signal) is reformed into a good sine wave, phase rotated by a resolver and then directed to the same phase detector. The phase detector develops a DC signal proportional to the phase difference with the difference being plus or minus accord-

ing to whether the reference signal leads or lags the second signal. This DC signal is chopped by a 400 cycle chopper amplifier and used to drive a servo motor. The resolver used to phase route the reference signal is driven by this servo motor. The net effect is that the shaft of the resolver repeats mechanically the elevation co-ordinate recorded on the magnetic tape. A second such servo repeats the azimuth co-ordinate and a third servo repeats the range co-ordinate.

The resolver shafts are marked as being geared up by 160 times, however, the ratio depends on how the shaft is compared with the corresponding radar shaft. In the case of the Nike System, this would actually be 250 times.

Each servo shaft is compared with the corresponding radar co-ordinate through a synchro control transformer which is geared down by a factor of 10 from the resolver shaft. The output of the synchro control transformer has a magnitude proportional to the difference or error between the radar co-ordinate and the dynamic tester co-ordinate and a phase according to the direction of the error. These errors are converted into DC errors which are plus or minus according to the direction of the errors.

These errors in azimuth, elevation and range are then handled differently according to the type of radar involved. For a monopulse system, a synchronizing pulse from the radar (sometimes called the anticipation pulse) is delayed a fixed period of time plus or minus a period which is proportional to the magnitude of DC range errors. This delayed pulse then turns on an oscillator for the duration of the pulse. The oscillator is tuned to the desired radar intermediate frequency. The IF pulse is attenuated according to range and sent to the radar range IF amplifier. The IF pulse is also sent to an azimuth phase and amplitude control circuit and an elevation phase and amplitude control circuit so as to convert the azimuth and elevation DC errors into IF pulses the radar will find normal. The control transformers, error demodulators and phase and amplitude control circuits simulate the radar antennas and microwave plumbing.

The system as described and shown in Figure 1 allows a program to be repeated any number of times with wear on the magnetic tape being the final limiting factor. Programs for storage on the magnetic tape can be developed in a number of ways but this paper is particularly concerned with storing a program that is generated by a command guidance system during a firing operation. The programs recording involves the dynamic tester used in a modified manner so that the tester servos follow the radar instead of the radar following the tester servos.

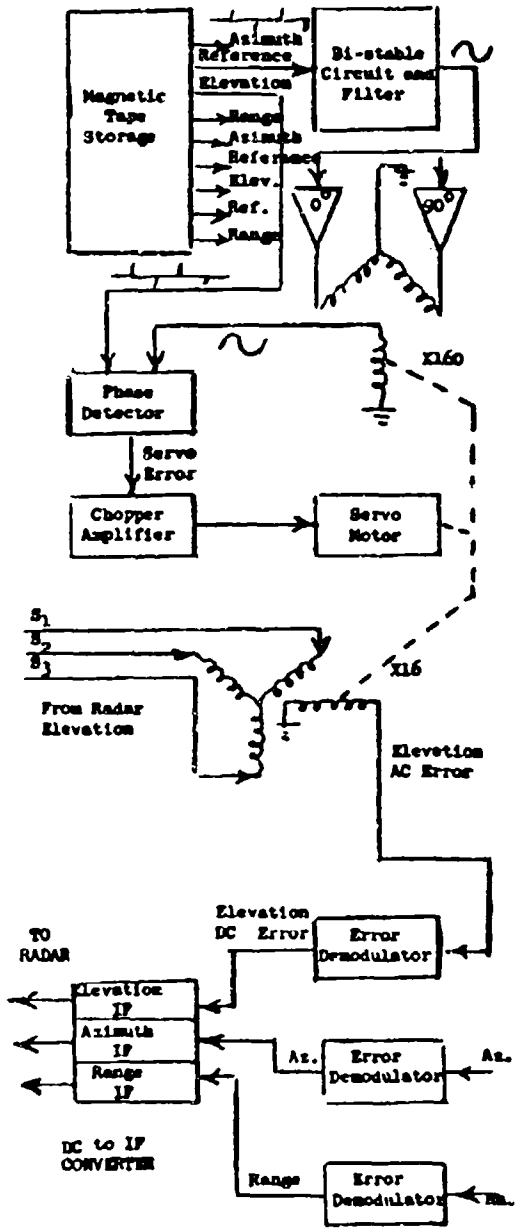


FIGURE 1

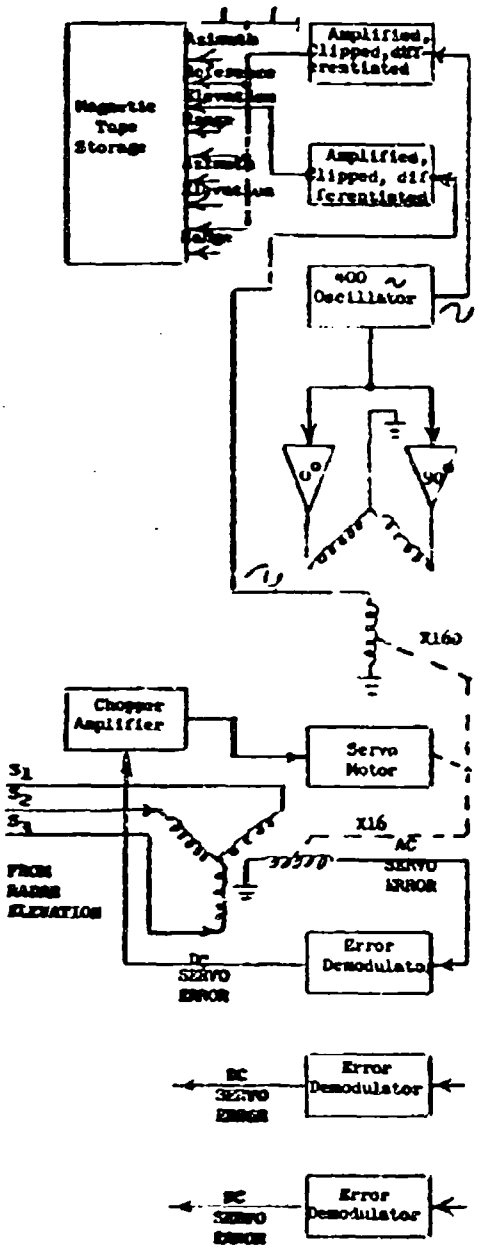


FIGURE 2

The program as described above is recorded on magnetic tape in analog form as phase comparison of 400 cycle per second reference and phase shifted signals. In the case of a Nike Hercules, a complete phase change of 360 degrees would represent a radar range change of 200 yards or an azimuth or elevation change of 25.6 mils. It can be seen that if errors in the system when referred to the program recording are less than 1 %, the program referred to the radar would be accurate to 2 yards or 0.256 mils or if the accuracy were 0.5%, the program referred to the radar would be accurate to 1 yard or 0.128 mils. The object of this test system is to keep errors to less than plus or minus 1 part in 200 as referred to the stored program on the magnetic tape.

This program would be stored on the tape as follows: (Figure 2)

(1) The three phase voltages (S_1, S_2, S_3) from the synchro transmitters on the X25 azimuth and elevation shafts and the 2,000 yard shafts are put through buffer amplifiers into synchro control transformers on the shafts marked "X16" in the servo system of the tester. The shaft ratio difference will not greatly matter since the X1 shaft of the tester is not used except for visual readout. The readout would of course be in error by the factor of $\frac{2}{16}$.

(2) The output of the control transformers go through buffer amplifiers to the phase detectors which are in the antenna simulation unit. The normal use of this unit is to convert these errors between the recorded program and radar into errors that are understandable to the radar and insert them into the front of the radar IF amplifiers. In this use of the antenna simulation, the errors are used to drive the servos in the tester. Only the phase detectors are used, which convert 400 cps errors to plus or minus DC errors.

(3) The net result is that the servos in the tester track the azimuth, elevation and range shafts of the radar. Noise which appears on the radar shafts will also appear on the tester servo shafts. The tester servo response should be adequate to maintain an accuracy of 1 part in 2560 between the X25 and 2,000 yard shafts and the corresponding tester servo shaft marked "X16".

(b) The resolver in the tester servo is not part of the servo system as described above but is part of the tester servo during playback operation. A reference 400 cycle per second sine wave is put into the resolver driver and also amplified, clipped and differentiated and recorded on magnetic tape. The phase shifted 400 cycle per second output of the resolver is also amplified, clipped and differentiated and recorded on an adjacent channel of magnetic tape. The phase relationship between the signals on the two channels of tape would be directly related to the azimuth, elevation or range of the radar. A 360 degree change in phase relationship would represent a 25.6 mil change in azimuth or elevation or a 200 yard change in range.

For test of the above procedure, the synchro (noted as radar servo) would be mounted on the shaft marked "X16" on a second tester servo which will be referred to as the control servo. This servo would be set to be positioned manually. Another servo, which will be referred to as the repeater servo, would be set for normal operation and aligned. The loop of the repeater servo would then be opened by opening the switch at the output of the servo pre-amplifier. A curve would be run, plotting actual mils error against the voltage at the output of the servo pre-amplifier. The repeater servo is next reconnected as described above so that the synchro control transformer instead of the resolver is in the servo loop. The voltage level at the output of the servo pre-amplifier is then adjusted to match the first curve by adjusting the level of the error detector output.

After completing the above checkout, the repeater servo will follow the position of the control servo. A curve should then be run to verify this and any errors noted. The loop of the control servo can now be closed and controlled by a program on magnetic tape or by a function generator. The output of the phase detectors in both servos should be then recorded on the error recorder in order to determine the servo errors.

Assuming that all errors are low or less than 1 part in 2560 at the synchro, the servos are now ready to be controlled by the radars or to control the radars as is desired. Six radar shaft positions can be recorded and repeated to the above accuracy. The program recorded can also be tabulated with time at a rate of 5/sec in order to determine the program to within 1 part in 64,000 as referred to absolute space position.

A time code reader is used to read a binary code from the magnetic tape record and provide the time to control a printer and a time pulse to control shaft angle digitizers so that shaft angle tabulations correspond to the correct time. This means that each time a certain recorded program is played back the azimuth, elevation and range tabulations are always the same for a certain time tabulation.

The digitizers in the case of the Nike Hercules are on the 256 mil azimuth and elevation shafts and on the 2000 yard range shaft. Since the digitizer breaks 360 degrees of the shaft angle into 4000 parts, 6400 mils for azimuth or elevation is broken down into 0.064 mil increments and range is broken down into 0.5 yard increments.

A direct writing recorder with the Melpar tester is designed to record the command guidance missile system performance data.

The tester as designed by Melpar Inc. with some re-routing of signals becomes a device to repeat up to 6 shaft positions from 2 guidance radars to an accuracy of 1 part in 64,000. The 400 cps voltages to the 6 resolvers on the tester shafts are recorded on magnetic tape along with the 400 cps voltages at the output of the resolvers. For optimum economy of tape channels, the same 400 cps signal can be sent to each resolver. This will allow 3 reference channels and 6 signal channels to represent 6 shaft positions with respect to time for a total of 9 magnetic tape channels. This leaves one channel for a time code and 4 channels for other data such as commands. Visual data requirements can be on the direct writing recorder which is part of the tester.

Once recorded the magnetic tape record and visual record can be sent to any like command guidance missile system and a direct comparison made. This is analogous to operating two systems, such as two Nike Hercules systems, side by side tracking the same target and the same missile and developing two sets of commands. One set of commands, commands the missile and the other set is compared to it.

The question that should now be asked is: Why is it necessary or important to certify the accuracy of the command guidance system? The test equipment, which is part of most command guidance systems, is quite adequate for normal checks, however, an area commander must have more assurance of the system accuracy than can be provided by the normal checkout or test equipment. He must have a counter-check or a means of sending in an evaluation crew with special equipment to certify that the command guidance system is completely ready and up to standard. The certification must be done rapidly, accurately and without shut-down or de-activation of the command guidance system. The dynamic test procedures as described above are proposed for this purpose.

SYSTEM EVALUATION OF LOW-NOISE RADAR SENSITIVITY

By: S. Charton and G. VerWys

INTRODUCTION

Radar system sensitivity is determined by the total noise in the receiving portion, including noise of both internal and external origin. Until fairly recently, the radar receiver noise level was usually so much higher than the level of natural external noise, that the environmental noise level was of secondary significance (barring intentional and unintentional man-made interference). However, recent reductions in receiver noise level by variable reactance amplifiers, masers, etc. have resulted in externally-originating and internally-originating noise levels of the same order of magnitude. Thus, a careful evaluation of the external noise environment must be made in order to determine system sensitivity.

The external noise arising from such sources as atmospheric oxygen and water vapor, the galaxy, radio stars, etc. can be readily estimated based on measurements made in the fields of radio astronomy and propagation. (For examples, see references 1, 2, and 4). Warm earth radiation reaches the antenna from the sidelobes and spillover, and at times even through the principal beam. If certain assumptions concerning terrain, antenna pattern, ground cover, etc. are made, earth noise can be estimated. However, little experimental data had been presented to justify these estimates. Therefore, it was decided to make experimental measurements on representative antennas and sites in order to check the validity of the theoretical estimates.

Accordingly, antenna noise temperature measurements were made for various pointing angles on both an 85-foot and 60-foot diameter parabolic UHF reflector and feed system, at two different locations. Measurement results were in substantial agreement with predicted results. It thus appears that it is possible to rapidly estimate the antenna noise temperature and hence system sensitivity, with an accuracy suitable for most system considerations.

NOISE TEMPERATURE TERMINOLOGY

It has become common practice to express noise generated in a circuit and the noise figure of a device in terms of noise temperature in degrees Kelvin. The advantage of using the noise temperature nomenclature is two-fold. First, consistent units can be used for comparing antenna noise with other noise generated in a radar system. Secondly, the overall radar sensitivity can be conveniently obtained, taking into account antenna noise contributions.

The noise figure of a two port transducer (see Figure 1) is expressed as:

$$F_1 = \frac{\text{available signal input power}}{\text{noise input power}} \Big/ \frac{\text{available signal input power}}{\text{noise output power}} = \frac{S_{in}}{N_{in}} \Big/ \frac{S_{out}}{N_{out}} = \frac{N_{out}}{G_1 N_{in}} \qquad (1)$$

The noise input, N_{in}, is just the Johnson noise generated in the input resistance.

$$N_{in} = k T_0 B \qquad (2)$$

where

- k = Boltzmann's constant
- T_0 = reference temperature = 290° Kelvin
- B = transducer bandwidth over which the noise figure and gain can be considered constant.

If N_1 = the noise generated within the transducer itself, then

$$N_{out} = N_{in} + N_1$$

$$N_{out} = k T_0 B G_1 + N_1$$

where G_1 = the gain of the transducer

If T_1 is the temperature of a resistance at the amplifier input which generates an amount of Johnson noise power equivalent to the amplifier output noise caused by the amplifier alone, then

$$N_{out} = k T_0 B G_1 + k T_1 B G_1 \qquad (3)$$

substituting (2) and (3) into (1)

$$F_1 = \frac{kT_0BG_1 + kT_1BG_1}{G_1 \quad kT_0B} = \frac{T_0 + T_1}{T_0}$$

Therefore the noise temperature of the transducer is

$$T_1 = (F_1 - 1)\, T_0 \qquad (4)$$

A complete theoretical analysis can be obtained from reference 8.

Carrying out the same operation for more than two stages in cascade yields:

$$T_{cascade} = (F_{cascade} - 1)\, T_0 = T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1 G_2} + \cdots + \frac{T_N}{G_1 G_2 \cdots G_n} \qquad (5)$$

where T_1, T_2, T_3, ... T_n and G_1, G_2, G_3 ..... G_n are the noise temperature and gain respectively of stages 1, 2, 3, ... n in the cascade network. $T_{cascade}$ and $F_{cascade}$ are the total noise temperature and noise

figure of the cascade network respectively.

Consider the case of a lossy transmission line at temperature T_p, with its input noise at a temperature T_A differing from T_0. (See Figure 2) The following expression gives the effective output noise temperature from such a network.

$$T_{eff} = \frac{T_0 + (L-1)T_p}{L} \qquad (6)$$

The utility of such an expression is seen when it is realized that T_A is any input noise temperature to the lossy network, and as such could be an antenna noise temperature. Using equation 6, an effective input noise temperature to a radar receiver, which accounts for the transmission line losses, can be found. A system noise temperature can be defined as the physical temperature of a resistor, which when placed directly across the input terminals of a noise free receiver, would generate the same amount of noise at the receiver output as the actual system generates. Combining equations 5 and 6, a system noise temperature can now be obtained. (See Figure 3)

$$T_{system} = \frac{T_A}{L} + \frac{T_p(L-1)}{L} + T_R, \qquad (7)$$

where T_R = noise temperature of receiver = $T_{cascade}$ of (5)

Using system noise temperature nomenclature, a figure of merit for the radar, (T_s), where the relative contribution of each system component can be readily deduced, is obtained.

SYSTEM NOISE TEMPERATURE

A. Antenna Noise Temperature

Upon closer examination of the noise-contributing components in a radar system, it is seen that a more detailed description of antenna noise temperature is required. Antenna noise temperature is the temperature at which the equivalent antenna radiation impedance must be maintained to generate an amount of Johnson noise equal to the noise power received by the antenna from its external surroundings. Antenna temperature then gives a measure of the amount of noise power received with the signal, and as such becomes a fundamental sensitivity limitation on any radar unit. Antenna noise temperature depends upon the direction of antenna pointing, and the objects or environs that are present in the actual antenna pattern path. This noise can originate from the black-body radiation of objects, or from other noise-generating processes. The magnitude of the black-body radiation is given by the Rayleigh-Jeans law and depends upon the absorption coefficient and temperature of the body in question.

The antenna noise temperature then is actually a representation of received noise power by an equivalent amount of Johnson noise power generated by the antenna radiation resistance at this physical temperature.

This received noise power may come from many sources, most of which are included in the following list.

1. Galactic background radiation;
2. Radiation from the earth into the antenna sidelobes;
3. Radiation from the sun, moon, and planets;
4. Radio star noise;
5. Atmospheric oxygen and water vapor;
6. Indirect or reflected radiation from the above sources, such as atmospheric scattering of warm earth radiation.

The relative amounts contributed to the total antenna temperature by the individual sources listed above will depend upon the antenna pattern, the pointing direction of the antenna, and the frequency of operation.

B. Transmission Line Losses

For high-power, large-aperture, mechanical tracking radar systems there usually are transmission losses between the antenna and the first amplification stage of the receiver. The main transmission losses occur in the transmission line itself, the rotary joints, the duplexer, and any couplers or monitoring devices added before the receiver. In a typical system, the transmission line loss could be 3/4 db, the rotary joint loss 1/4 db, and the duplexer loss 3/4 db, giving a total transmission loss of 1 3/4 db. Figure 4 shows the effective antenna temperature seen by the receiver as a function of loss between antenna and receiver. T_{eff} is given by equation 6. For instance, if T_A were 100° K, and there was a 2-db line loss, the receiver would see 170° K. If T_A were 0° K, and there was a 2-db line loss, the receiver would see 106° K.

C. Typical Receiver Noise Temperatures

The receiver noise temperatures to be expected in the UHF region of interest depends on the type of front-end amplifier used. A conventional UHF tube such as the 416-B can be expected to give a noise figure of approximately 5 db or 630° K noise temperature. Using a parametric amplifier, a noise figure of 1 db or a noise temperature of 75° K could be attained. Most other devices such as crystal mixers would result in a higher noise temperature than that of UHF triodes. It can be seen that with the use of a parametric amplifier, the antenna noise temperature becomes very significant in any system evaluation, because the antenna noise temperature becomes comparable in magnitude to receiver noise temperature.

ANTENNA NOISE TEMPERATURE MEASUREMENTS

A large amount of data have been compiled by radio astronomers concerning sky temperatures and radio-

wise noise. These data have generally been taken with special effort to minimize antenna backlobe and sidelobe effects. The actual noise temperature of a radar antenna, considering sidelobes and spillover, must be predicted on the basis of the particular antenna configuration and geographical location. Many articles have been published on methods of computing antenna temperatures, but few experimental data are available to verify the computational methods at the frequency of interest. Therefore, measurements were made on two large-aperture antennas to verify expected antenna temperatures. The first set of data was taken on the Millstone Hill tracking radar operated by Lincoln Laboratories. The second set of measurements was made on a Stanford Research Institute radar located in Fairbanks, Alaska. Results of these measurements are presented in the following paragraphs.

A. Antenna Temperature Measurements at Millstone Hill

A radiometer built by Airborne Instruments Laboratories was used to make measurements of antenna temperature versus antenna azimuth and elevation. The radiometer was a switched, self-balancing type which provided a continuous recorded output of the noise temperature at its input terminals. The radiometer scale was calibrated at five different temperatures. A termination, immersed in liquid nitrogen was used for calibration of three scale points by adding three different amounts of attenuation between the termination and the radiometer terminals. A room-temperature load and a load immersed in boiling water constituted the other two calibration points. Since the radiometer was calibrated at its input terminals, the indicated antenna temperature readings had to be corrected for the transmission line loss. A block diagram of the test facility is shown in Figure 5.

Data were taken starting at 12:00 midnight and ending at 12:00 noon the next day. Azimuth sweeps of 360° were made for constant elevation angles of 0°, 2°, 4°, 6°, 7°, 8°, 10°, 15°, 20°, 30°, 45°, 60°, and 90°. The time required for a 360° azimuth sweep was 15 minutes, and an integration time constant of 4.5 seconds was used for all measurements. With a 2° beamwidth, approximately 15 minutes were needed to allow one integration time constant per beamwidth of azimuth sweep. The outside physical temperature was between 25°F and 35°F, and a four-to six-inch snow cover existed. The Millstone radar antenna feed system included two orthogonally-polarized receiver channels. The results quoted here are for one channel.

Figure 6 shows a reproduction of one azimuth sweep. The average quiet level of this constant 7° elevation sweep is indicated at scale reading 375, which corresponds to an antenna temperature of 60° K. Clearly shown are two peaks which correspond to the Milky Way (smaller peak) and the sun. Figure 7 shows a plot of average quiet level antenna temperatures over an azimuth sweep versus elevation angle. This curve indicates that the antenna temperature does not vary appreciably after the main beam and principle sidelobes are pointing above the ground, providing no hot galactic sources are present within the beam. After 15° elevation, there is only a gradual decline in antenna temperature at increasing elevation. This gradual decrease is presumably at least partially caused by decreasing atmospheric path lengths. During the morning hours it was noticed that, when the antenna is within 5° of the sums position, a noise temperature greater than 200° K existed. There seemed to be good correlation between tabulated sky noise sources and the indicated levels from these measurements. The Milky Way was always observed at the proper azimuth and elevation, and exhibited approximately the proper shape. A tabulation was made of all interference peaks encountered and the azimuth sector and time in which they were encountered.

B. Antenna Temperature Measurements at College Alaska

The antenna used in these measurements was a 60-foot parabolic reflector with a dipole feed system. This antenna was made available by Stanford Research Institute. The nominal beamwidth of the antenna was 3°.

The radiometer consisted of a very stable Ewen Knight LNRE 400-mc receiver whose detected output was displayed on a DC chart recorder. This receiver was made available by Lincoln Laboratories. Figure 8 shows the test facility block diagram. The calibration was accomplished by placing a matched load in liquid nitrogen and using several different attenuation values between the load and the calibration point at the receiver input terminals. This calibration was checked every 10 minutes during the recording time, since this type of radiometer is sensitive to drifts in amplifier gain.

Data were taken for the full 360° azimuth sweep and elevation angles of 0°, 2°, 4°, 6°, 8°, 10°, 20°. The time constant was adjusted to 7 seconds, with azimuth scan rate of 20 minutes for 360°. Figure 9 shows a plot of elevation angle versus average quiet antenna temperature.

ANTENNA NOISE TEMPERATURE PREDICTION

Antenna noise temperature, as mentioned earlier is simply a constant time the noise power received by an antenna from its environment. This received noise power is the integral of the product of antenna gain and the noise power radiated by the environment. Thus the following expression can be used as an alternative definition of antenna noise temperature:

$$T_A = \frac{\int_S T(\theta, \phi)\, G(\theta, \phi)\, dS}{\int_S G(\theta, \phi)\, dS}.$$

where $T(\theta, \phi)$ = noise temperature associated with direction (θ, ϕ)
$G(\theta, \phi)$ = antenna gain in direction (θ, ϕ)

Thus, writing the above expression in discrete form (see Figure 10), one obtains

$$T_A = \sum_i a_i T_i,$$

where $a_i = G(\omega) \frac{\omega}{\beta s}$.

and $G(\omega)$ = average gain over solid angle ω, and T_i is average noise temperature associated with solid angle s.

It can be seen that a_i is a weighting factor, equalling the fractional portion of power entering the antenna in any given solid angle. Thus, antenna noise temperature equals the weighted sum of the noise temperature contributions in the antenna pattern.

As an illustration, the antenna noise temperature at an elevation angle of 5° will be calculated. A simplified diagram of the antenna pattern is shown in Figure 11a. A representative value for the fractional part of the power received from the main beam is 0.8. The remainder of the received power will, for simplicity, be assumed to be uniformly distributed throughout the total solid angle, with half of this power coming from below 0° elevation, and the remainder from above 0° elevation.

The table shown in Figure 11b indicates the method of calculating this antenna noise temperature.

The oxygen and water vapor in the atmosphere, which cause r-f attenuation, also generate r-f noise. Since the path length through the troposphere varies with elevation angle, the associated resultant noise temperature also varies with elevation angle. From previously-measured values of attenuation as a function of frequency and elevation angle, the resultant noise temperature can be calculated. These values have been tabulated and plotted by several investigators. (References 1, 2, and 4.)

At the frequency of approximately 400 mc, at which radiometer measurements were carried out, the antenna temperature due to oxygen and water vapor absorption at 5° elevation angle was 12° K. The weighting factor used was 0.8, that associated with the main beam. Thus, the contribution due to the above factor was 10° K. (Note that the contribution due to the above factor in the sidelobes is a second-order effect, because most of the power entering the antenna pattern comes from higher elevation angles, where the atmospheric path is relatively short and the resultant noise temperature low; and secondly, because of the 0.1 weighting factor.)

The noise temperature contribution due to thermal radiation from the earth acting as a black body was 270°K, the physical earth temperature when the radiometer measurements were made (February, 1959). With a weighting factor of 0.1, due to this noise entering only into the lower sidelobe hemisphere, the resultant antenna temperature due to earth radiation was 27°K.

The contribution due to galactic noise can be readily found by examination of galactic noise charts such as are found in References 5, 6 and 7. The cosmic noise intensity varies over approximately a ten-to-one range at a given frequency, depending on whether one is pointed at the hot part of the Milky Way or in a more noise-free portion of the sky. The average value of cosmic noise temperature at 400 mc in the noise free portion of the sky was estimated at about 20°K. With a weighting factor of 0.9 due to the main beam and sidelobes in the upper hemisphere, the resultant cosmic noise contribution was 18°K.

Inasmuch as the radiometer measurements plotted were those made at night, it was not necessary to account for contributions due to sun noise. Contributions from radio star sources were negligible, as their noise arrived at the low sidelobe levels. Summing up the important noise contributions, as shown in the table of Figure 11 a net antenna temperature of 55°K results. The measured antenna noise temperature for this 5° elevation angle was 60°K at the Millstone Hill radar, and 57°K at the SRI Alaskan radar. (Thus correlation between estimated and measured results is judged adequate for most system purposes). Calculation of estimated antenna temperatures at other elevation angles yielded results adequately close to measured minimum values.

The net conclusion is that fairly good estimates of antenna noise temperatures can be made, without the necessity for radiometer measurements, for most conventional antennas, with an accuracy sufficient for most system planning.

As an illustration of the use of antenna noise temperatures in calculating system sensitivity, the following example is given.

Assume two radar systems are to be compared. System I uses a receiver having a noise figure of 3 db, whereas System II uses a receiver having a noise figure of 1 db. System losses between antenna and receiver input are estimated as 2 db. Antenna noise temperature is assumed to be 100°K. What are the resultant system sensitivities, and what is the improvement in sensitivity of System I over System II? Figure 12 shows the calculation.

If reception losses were reduced from 2 db to 0 db, Figure 13a shows that the following sensitivities are obtained:

System I: $T_s = 100 + 290 = 390°$ K.
System II: $T_s = 100 + 75 = 175°$ K.

$$\frac{T_{s_I}}{T_{s_{II}}} = \frac{390}{175} = 2.23 \Longrightarrow 3.5 \text{ db.}$$

Now if T_A were reduced to 0°K, Figure 13b shows that the resultant system sensitivities are obtained:

System I: $T_s = T_r = 290°K$
System II: $T_s = T_r = 75°K$

$$\frac{T_{s_I}}{T_{s_{II}}} = \frac{290}{75} = 3.86 \Longrightarrow 5.9 \text{ db.}$$

Figure 14 is a plot of system noise temperature against receiver noise temperature for an assumed antenna noise temperature of 100°K. Notice the increase in system noise temperature of about 40°K for a 1-db line loss, and about 70°K for a 2-db line loss. This factor is most important for low values of receiver noise temperature.

SUMMARY

The original system, replacing a 3-db noise figure receiver with a 1-db noise figure receiver, yields a 2.7 db improvement in sensitivity. If there were no reception line losses, the improvement would be 3.5 db. Furthermore if there were zero antenna noise temperature, the improvement in sensitivity would be 5.9 db, due to improvement in receiver noise temperature. Note that receiver noise figure implies an antenna noise temperature of 290°K. Thus, replacing a 3-db noise figure receiver with a 1-db noise figure receiver yields an improvement in sensitivity better than 2 db whenever the antenna noise temperature is less than 290°K.

The above example points out the need for concentrating on all elements of a low-noise radar system — the antenna noise temperature, reception line losses and receiver proper. Present technology is such that the above contributions are of the same order of magnitude. For example, Bell Telephone Laboratories (Reference 3) has measured a 17.6°K system noise temperature of a low-noise maser receiving system at 5.65 Gmc frequency, broken down as follows:

Direct sky noise (at zenith)	2.5°K
Side-or Backlobe Pickup	2°K
Traveling wave maser	10.5°K
Antenna, waveguide and coupler loss (0.05 db)	3.5°K

This points out the trend of low-noise systems in the future. It is realized that it will be difficult for present radar systems to attain such low sensitivities when a single antenna is used for transmitting and receiving, due to unavoidable T-R tube losses, and physical location of the r-f head. However, if maximum sensitivity is to be obtained, the design implications are clear.

CONCLUSIONS

a. Antenna noise temperature estimates can be made rapidly without the need for radiometer measurements, for most system sensitivity analyses.

b. A low noise antenna, which minimizes radiation reaching it through the earth via spillover and sidelobes, is an important part of a low-noise receiving system.

c. It is important to minimize losses between antenna and receiver. Placing the receiver near the antenna feedhorn aids in cutting down losses. In addition, separate transmitting-receiving antennas eliminates duplexer losses and possible TR tube noise generation.

d. However, man-made interference poses a problem. Site interference surveys must be made and remedial action taken, if necessary, in order to benefit from a low-noise system.

e. In the limit, it can be said that the environment will put a bound on ultimate sensitivity of a radar system. Sensitivity will vary as the radar beam points to regions varying in noise level. For instance, radio stars, the sun, the moon, etc., will change the sensitivity of a system. System design may have to program the scan, for example, to avoid deterioration in sensitivity, or else will have to consider the varying sensitivity.

ACKNOWLEDGEMENTS

Acknowledgement is made of the efforts of the following individuals who have materially contributed to the measurement program.

RCA: E. W. Matthews and D. C. Venters

Lincoln Labs: L. Bird, L. G. Kraft, and A. Parsons

Stanford Research Institute: L. Kain and R. Bollen

REFERENCES

1. D. C. Hogg, "Effective Antenna Temperatures Due to Oxygen and Water Vapor in the Atmosphere," J. Appl. Phys., 30, 1417 (1959).

2. D. C. Hogg, W. W. Mumford, "The Effective Noise Temperature of the Sky," Microwave Journal, March 1960, p. 80.

3. R. W. DeGrasse, D. C. Hogg, E. A. Ohm, H. E. D. Scovil, Journal of Applied Physics, Vol. 30, p. 2013, Dec. 1959.

REFERENCES (Cont'd)

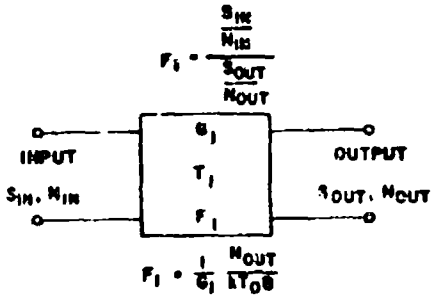
4. R. Gardner, "Final Engineering Report on Antenna Noise Temperature Study," AIL Report No. 3304-11, Nov. 1957. (Performed under Contract DA-49-170-SC-1547.)

5. N. G. Roman and S. S. Yaplee, "Radio Sources and the Milky Way at 440 MC," Proceedings of the IRE, Vol. 46, No. 1, January 1958, p. 199.

6. K. C. Ko, "The Distribution of Cosmic Radio Background Radiation," Proc. IRE, Vol. 46, No.1, January 1958, p. 208.

7. D. H. Menzel, "Cosmic Noise Survey," Harvard College Observatory, Cambridge, Mass.

8. Donald B. Harris, "Microwave Radiometry", Microwave Journal, April 1960, Vol. 3, No. 4.



$$F_1 = \frac{\frac{S_{IN}}{N_{IN}}}{\frac{S_{OUT}}{N_{OUT}}}$$

$$F_1 = \frac{1}{G_1}\frac{N_{OUT}}{kT_0B}$$

$$N_{OUT} = kT_0BG_1 + kT_1BG_1$$

$$F_1 = \frac{1}{G_1}\frac{kT_0BG_1 + kT_1BG_1}{kT_0B} = \frac{T_0 + T_1}{T_0}$$

$$T_1 = (F_1 - 1)\,T_0$$

Figure 1. Receiver Noise Temperature

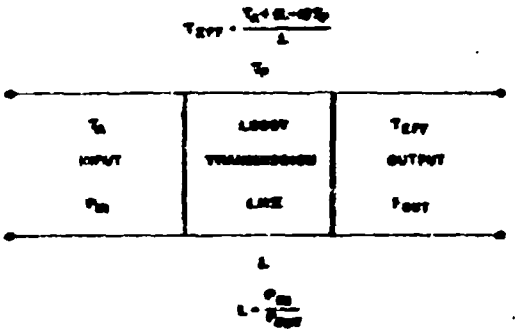


Figure 2. Effective Noise Temperature



$$T_S = \frac{T_A}{L} + \frac{(L-1)T_p}{L} + T_R$$

Figure 3. System Noise Temperature

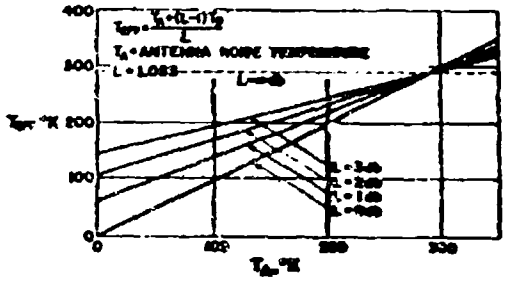


Figure 4. Effective Noise Temperature vs Antenna Noise Temperature

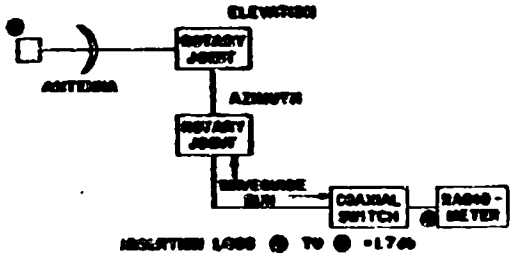


Figure 5. Millstone Test Facility

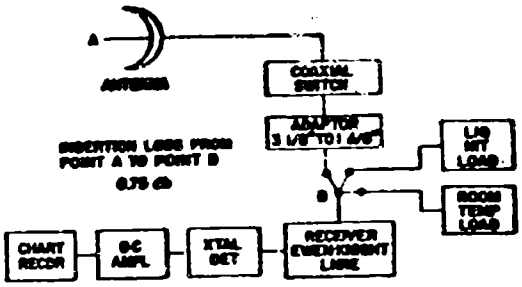


Figure 8. Block Diagram of Test Facility

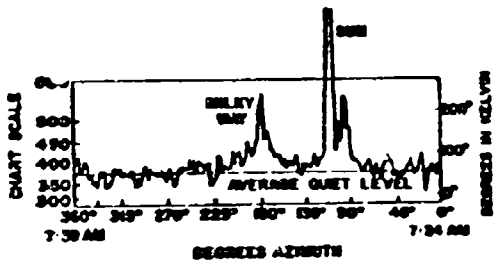


Figure 6. Typical Noise Temperature Recording Elevation 7 Average Quiet Level-50° K

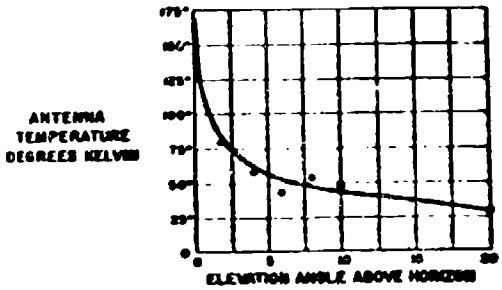


Figure 9. Antenna Temperature vs Elevation Angle SRI Alaska Radar

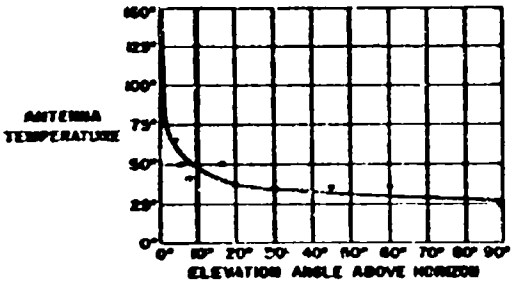


Figure 7. Antenna Temperature vs Elevation Angle Millstone Hill Radar

$$T_A = a_1 T_1 + a_2 T_2 \text{ ——— } + a_i T_i$$

WHERE

T_0 = ANTENNA NOISE TEMPERATURE

a_i = ANTENNA GAIN WEIGHTING FACTOR, FRACTIONAL PORTION OF POWER ENTERING ANTENNA

T_i = NOISE TEMPERATURE ASSOCIATED WITH SOLID ANGLE FOR WHICH a_i APPLIES

Figure 10. Antenna Temperature Calculations



CALCULATION OF ANTENNA NOISE TEMPERATURE AT ELEVATION ANGLE = 0°

CONTRIBUTION	AMOUNT, T_i, °K	WEIGHTING FACTOR, a_i	WEIGHTED TEMPERATURE, °K
ATMOSPHERIC ABSORPTION	12	0.8	10
EARTH	270	0.1	27
GALACTIC NOISE	20	0.8	16
ANTENNA NOISE TEMPERATURE			50°K

Figure 11. Antenna Temperature Calculation Example

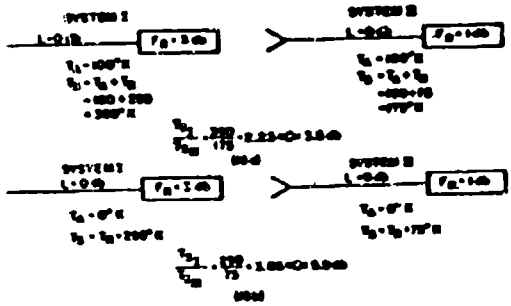


Figure 13. System Temperature Calculation Examples

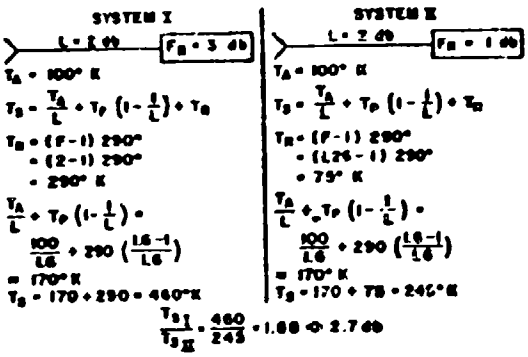


Figure 12. System Temperature Calculation Examples

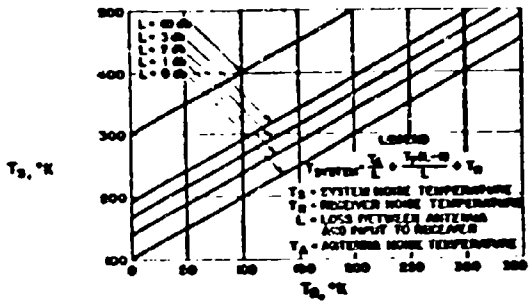


Figure 14. System Noise Temperature vs Receiver Noise Temperature for Antenna Noise Temperature = 100° K

A FLUSH-MOUNTED VHF TELEMETRY ANTENNA WITH HEMISPHERICAL COVERAGE

By Ronald C. Payne and Parker Painter, Dynatronics, Inc.

REQUIREMENTS

The requirements of this antenna as were outlined in the applicable technical specification were that the antenna be able to receive missile signals while the missile is in the upper hemisphere of the aircraft. Thus the antenna must present circular polarization characteristics in all directions above the horizontal plane of the aircraft or be circularly polarized directly overhead and vertically polarized at the horizon. This would permit reception of missile signals at any missile azimuth angle while in level flight and the vertical polarization in the horizon would permit reception of signals while the missile is in its terminal phases of flight. The axial ratio directly overhead must not exceed 1.26:1 and the VSWR must not exceed 2:1 over the 215 to 260 mc telemetry band. Nulls in the radiated pattern must not exceed 10 db. The antenna must present a minimum of aerodynamic drag and produce practically no air flow turbulence in the vicinity of the vertical and horizontal stabilizers. The antenna shall be suitable for installation in the upper fuselage of a C-54 aircraft.

For aerodynamic considerations, the above requirements suggest a flush mounting antenna, such as a slot antenna. Two slot antennas which are oriented 90° with respect to each other and phased properly, will give broadband operation and will exhibit circular polarization on an axis perpendicular to the plane of the slots and linear polarization in the plane of the slots. The direction of polarization in the plane of the slots will be perpendicular to this plane. A flush mounting slot antenna system offers no aerodynamic drag and no air flow turbulence; an extra feature of flush mounting is that precipitation static is greatly reduced over a bubble radome installation.

SLOT ANTENNA THEORY

PATTERN AND POLARIZATION

The radiation characteristics of a slot antenna in an infinite flat sheet are compared to that of a complimentary dipole by reference to figure 1. (Ref. 1, Pages 357, 358).

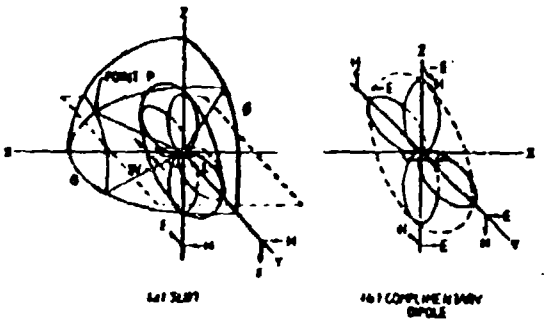


Figure 1.

The infinite flat conducting sheet is coincident with the x-y plane, and the long dimension of the slot is in the x direction. The complimentary dipole is coincident with the x axis. The radiation field pattern of the slot antenna has a doughnut shape identical to that of the complimentary dipole, except that the electric field vectors E and magnetic field vectors H are interchanged.

It is seen that the radiation in the x-y plane from the horizontal slot has a polarization parallel to the z axis, or is vertically polarized, and that along the z axis it is polarized parallel to the x-y plane or horizontally polarized. If the slot is very thin (the width is much shorter than the wavelength) and 1/2 wavelength in length, the variation of E_ϕ as a function of θ is determined by the equation:

$$E_\phi(\theta) = \frac{\cos(\pi/2 \cos\theta)}{\sin\theta}$$

For arbitrarily selected values of θ at 0°, 45°, and 90°, $E_\phi(\theta) = 0$, 0.62 and 1.0 respectively. If the sheet is infinite in extent and perfectly conductive, E_ϕ is a function of θ only (independent of ϕ).

Carrying this analysis further, if the sheet has some finite length (in the order of several wavelengths) a marked change in the $E_\phi(\theta)$ patterns results as seen in Figure 2 (a). (Ref. 1. page 359).

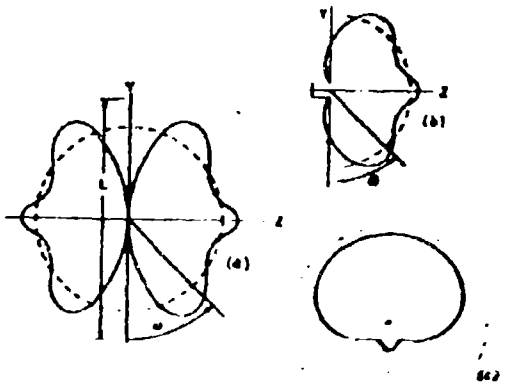


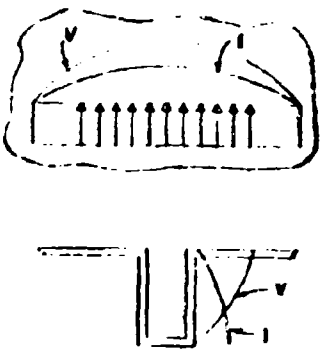
Figure 2.

A null occurs along the y axis (perpendicular to the axis of the slot) due to oppositely phased fields of equal magnitude being radiated from the two sides of the sheet. However, if one side of the sheet is boxed in, the field strength pattern appears as the solid line in Figure 2 (b). As the dimensions of the sheet are increased toward infinity a semicircular pattern is approached as shown by the dotted line in this figure.

If a slot is placed in a cylinder, there is no longer a null along the y axis since the ground plane is now continuous and there is no outward radiation from the inner side of the slot to produce cancellation. The radiation pattern for a slot in a cylinder whose diameter to wavelength ratio is 2.5 (which is the case for a C-54 aircraft at 240 mc) is shown in Figure 2 (c).

IMPEDANCE (VSWR) AND BANDWIDTH

The center point impedance of an open half wave slot with a relatively high length-to-width ratio is approximately 500 ohms. When the slot is boxed in on one side so that zero susceptance is shunted across the slot terminals, the impedance is doubled or becomes 1000 ohms. Since the current and voltage relationships in a slot radiator are as pictured in Figure 3 (Ref. 2, Page 713), it follows that a low impedance point can be achieved by exciting the boxed slot near the bottom of the cavity and toward one end where the voltage (and impedance) is low. However, the selection of this point and probe configuration are critical for broadband operation since reactive impedances appear as one moves away from the center of the slot. Also, the depth of the slot has an important bearing on the reactive component of the driving point impedance.



SLOT CURRENT VOLTAGE DISTRIBUTION

Figure 3.

In practice, the exact location of this point and slot depth is selected by experimentation to obtain the optimum bandwidth and impedance match to the transmission line.

As with dipole radiators, the broadbanding of the slot radiator is a function of the length-to-width ratio of the slot, the bandwidth increasing with lower ratios. Maintenance of low VSWR over a wide frequency range is a function of the cavity dimensions primarily the length to width ratio, and the method of feeding.

If there are two slots, crossed in such a fashion so that they will be mutually perpendicular, neither slot may be conveniently excited from the center without the possibility of exciting the other. As will be shown later cross coupling will cause difficulty in radiating a circularly polarized wave and can seriously limit bandwidth. It has been shown (Ref. 3, Pages 182, 183) that the approximate driving point impedance of a slot radiator decreases as the square of the distance away from the center of the slot and the square of the length of the probe. The driving point impedance also decreases as the probe is moved toward the bottom of the slot. Moving the driving point away from the center of the slot also decreases the bandwidth of the slot radiator to a slight extent. Driving probe diameter al-

so affects bandwidth. A thin cylindrical probe imparts a high reactance to the driving point impedance while larger diameter probes provide more coupling and have less self reactance.

EXPERIMENTAL RESULTS WITH SINGLE SLOT

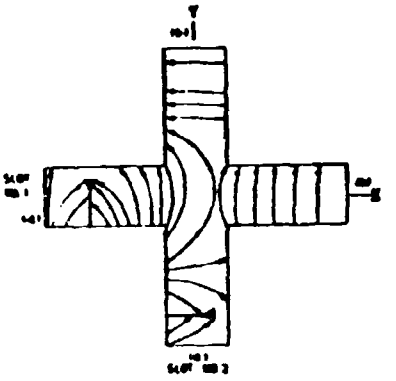
It was found experimentally that all of the aforementioned properties of a slot radiator are true. Additional information obtained was the driving point characteristics as the driving point was moved toward the bottom of the slot cavity. Changing the cavity depth has primary effect on the susceptance along the wall of the slot and in turn the reactive component of driving point impedance changing the driving point position perpendicular to the plane of the slot will change both the reactive and resistance components of driving point impedance. Therefore, by choosing a particular probe diameter and length, at a certain location and with a proper cavity depth, almost any driving point impedance can be realized. Using large cylindrical probes which were tapered on one end to provide a smooth transition to the coaxial feed system, a driving point was found which could easily give a 40% bandwidth for a VSWR of 2:1 for each of 50 ohms and 100 ohms.

CHARACTERISTICS OF CROSSED SLOT ANTENNAS

CROSSED SLOT RADIATION THEORY

As can be shown by reciprocity, the r-f characteristics of an antenna are identical whether transmitting or receiving, therefore, the crossed slot array is considered a transmitting antenna in the following analysis in order to facilitate discussion.

It is seen from Figure 4 that the electric field vector E bridges the crossover point by a radial distribution of energy much the same as that emanating into space from the slot itself. Since the electrical fields from the respective slots are mutually perpendicular, there is no cross coupling between slots assuming each slot is driven symmetrically. The magnetic field component H generated by the ground plane currents is distributed along the length of the slots. The ideal field strength patterns generated by a crossed slot arrangement (antenna installed on a flat, perfectly conducting sheet of infinite length) are two mutually perpendicular holeless doughnut patterns with their field vectors phased 90° in time.



THE CAUSE OF CROSS-COUPLING BETWEEN SLOTS
Figure 4.

In the actual installation, the slots are oriented 45° with respect to the longitudinal axis of the aircraft. This arrangement provides good electrical balance between slots, assures uniform overall coverage, and exhibits almost unity axial ratio directly overhead where it is most important. A close approximation of the pattern which can be expected from slots oriented in this manner is about half-way between the patterns obtained from slots (1) parallel with and (2) perpendicular to the longitudinal axis. In the first case, a considerable length of fuselage serves as a ground plane and, except for a slight modifying effect due to sloping off of the fuselage, the pattern approaches that represented by the dotted line in Figure 2 (b). For Case No. 2, the slot has as its ground plane, in effect, a cylinder whose diameter is that of the aircraft. The radiation pattern from a slotted cylinder with a diameter to wavelength ratio of 2.5 is approximately that shown in Figure 2 (c) (Ref. 4 page 689).

The two antennas provide patterns about the z axis which are identical and are oriented with respect to each other at an angle of 90°. Further, the electric vectors at each point in space are at angles varying between 90° on the z axis and 0° in the x-y plane, but always time phased 90° apart. In analyzing this vector arrangement two axes u and v are defined, with the u-v plane perpendicular to the radius vector from the antenna, and the u axis always parallel to the x-y plane (see Figure 5).

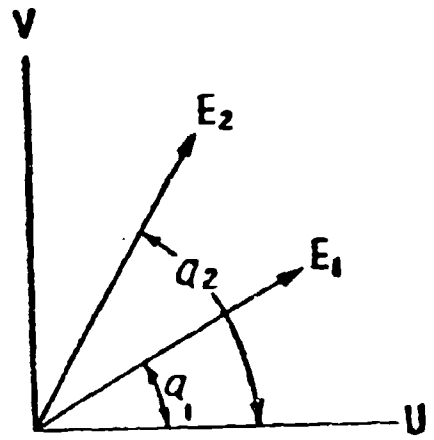


Figure 5.

The electric vectors, E_1 and E_2 from the antenna will lie in this u-v plane. Therefore:

$$E_v = E_1 (\sin a_1 \pm E_2 \sin a_2) \sin wt$$

$$E_u = E_1 (\cos a_1 \pm E_2 \cos a_2) \sin (wt+90^\circ)$$

$$= E_1 (\cos a_1 \pm E_2 \cos a_2) \cos wt$$

Let

$$K_v = E_1 \sin a_1 \pm E_2 \sin a_2$$

and

$$K_u = E_1 \cos a_1 \pm E_2 \cos a_2$$

then

$$E_v = K_v \sin wt$$

$$E_u = K_u \cos wt$$

which are the parametric equations for an elliptically polarized wave (of axial ratio K_v/K_u) at a point in space. For example, at a point directly overhead the magnitudes $E_1\phi(0)$ are the same for each slot.

Since the u and v axes are parallel to the x and y axes, respectively, (still directly overhead),

$$E_u = E_1\phi = 1.0 \cos wt$$

$$E_v = E_2\phi = 1.0 \sin wt$$

and the wave is circularly polarized (axial ratio A.R. = 1.0)

At an elevation of 60° in the y-z plane

$$E_1\phi(\theta) = \frac{\cos (90^\circ \cos 60^\circ)}{\sin 60^\circ} = .815$$

$$E_2\phi(\theta) = 1.0$$

$E_1\phi(\theta)$ is parallel to the u axis,

$\cos a_1 = 1$, and $\sin a_1 = 0$.

$E_2(\theta)$ is parallel to the v axis,

$\cos a_2 = 0$, and $\sin a_2 = 1$

$$K_v = 0 \pm E_2 \sin a_2 = E_2\phi = 1$$

$$K_u = E_1 \cos a_1 + 0 = E_1\phi = 0.815$$

Therefore,

A.R. (axial ratio) = $K_v/K_u = 1/0.815 = 1.22 = 1.72$ db. Similarly, the ellipticities at 45°, 30° and 0° are 4.1 db, 7.6 db and infinity, respectively.

The antenna arrangement will then provide excellent circularly-polarized vertical coverage when the missile is passing overhead in horizontal flight, and will continue to give good circularly-polarized coverage down to about 45° from the vertical axis. Beyond this point, although the ellipticity increases to infinity, full vertically-polarized coverage is provided during the diving phase when the missile is normally emitting vertically-polarized waves.

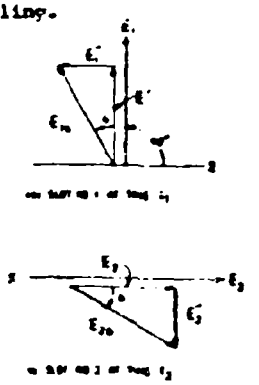
IMPEDANCE (VSWR), BANDWIDTH, AND CROSS COUPLING

A crossed-slot antenna was constructed to the dimensions stated earlier as optimum for a single slot and placed in a flat ground plane. The first effects noted were a decrease in bandwidth and changes in input impedance. A new problem associated with driving crossed-slots is that, when the slots are phased 90° electrically, the driving impedances must be such that the power division between the two slots will be nearly equal in order to achieve good circular polarization directly overhead.

In testing to improve VSWR, it was reasoned that a slight dielectric loading at the mouth of the slot introduces enough additional capacitive reactance at the slot opening to permit the physical depth of the slot to be decreased. By utilizing a 5/16 inch thickness of epoxy fiberglass

as a radome, the depth of the slot may be shortened a full inch. This is reasonable since the relative dielectric of epoxy fiberglass is such that the 5/16 inch epoxy produces a total dielectric equal to approximately one inch of air. This dielectric is effective only near the mouth of the slot where the electric potential is greatest. Crossing the slots cause one other serious degradation to the system, a high degree of cross-coupling between slots. If power is entering the system by way of the driving point in Slot #1 (See Figure 4), energy radiates from the mouth of the slot and also into the region common to both slots.

The fringing of the electric vectors, as shown in Figure 4, causes energy to be coupled to both ends of slot No. 2, and the probe in that slot couples some of this energy back to the feed system. This indicates that the feed system should include a hybrid ring to provide a high degree of external isolation between slots. This would help eliminate the effect of cross-coupling on VSWR, but not to a large extent, the effect being imposed upon the antenna radiation pattern. As shown in Figure 6 the electric vectors at times t_1 and t_2 (where t_2 is 90° later than t_1) are not spaced-phased 90° under conditions of cross-coupling.



EFFECT OF CROSS-COUPLING ON ELECTRIC VECTORS

Figure 6.

Consider a time t_1 when the input voltage is maximum on slot #1 and zero into slot No. 2. E_1 in Figure 6 (a) is the voltage vector radiated from both ends of slot #1, neglecting the effect of cross-coupling. Since some power is being coupled into slot No. 2 as in Figure 4, E_1 does not have the same magnitude in the y direction; instead there is a smaller vector E_1'. This is caused by some of the power being coupled into Slot No. 2 and being fed back to the feed system at the probe end of this slot. Assuming a perfect match in the (a) end of Slot No. 2 between the probe and its coaxial termination, all of this power is dissipated. Therefore, the (a) end of Slot No. 2 has a vector in the plus x direction and the (b) end of Slot No. 2 has a larger vector in the minus x direction. The resultant is a component of power which enters Slot No. 2 at (a), appearing in the minus x direction which is indicated in E_1'' in Figure 6 (a). The resultant is a vector E_{1b} with a space phase of 90° ± b. By the same analysis, when time has advanced to t_2 (90° later in time) the radiated vector from Slot No. 2 is E_{2b} at an angle of -6 degrees. Therefore, the total space displacement for the electric vectors from each slot is 90° ± 2 b, which is effectively the physical angular displacement between slots. Therefore, it is seen that good circular polarization is difficult to obtain with this arrangement. It would be possible to electrically phase the two slots in the feed harness by an amount equal to 2 b and maintain good axial ratio over a narrow range of frequencies, but not under broadband conditions. Laboratory tests indicated that for the best compromise in electrical phasing of the feed harness, ellipticities at certain frequencies were in excess of 6 db.

Even though it was possible to correct the ellipticity to a tolerable degree by electrical phasing, the horizontal pattern suffers considerably under these conditions of cross-coupling. Consider again Figure 4. It is seen that the driven end of the slot radiates more power than does the other end. Also, the energy which has been cross-coupled to the other slot is not radiated equally between the two ends. Therefore, the effective phase center has shifted away from the geometric center of each slot radiator.

ELIMINATING CROSS-COUPLING

At this point the objectives were to redesign the probe configuration to provide minimum probe fringing, to place the probe toward the end of its slot that the probe fringing is smoothed out before reaching the center of each slot, and to provide some means to cancel out the power which was being coupled to the adjacent slot.

A sharply tapered probe made of a small rod and fins was developed, which provides

a much tighter coupling and causes less fringing around the probe area. A probe was placed in both ends of each slot, and the probes of a given slot are driven in phase. Thus, energy coupled to Slot No. 2 from slot at (a) is cancelled by the energy coupled into Slot No. 2 from Slot No. 1 at (b). With this arrangement, the degree of cross-coupling was not measurable even with highly sensitive impedence measuring devices. The technique used in determining cross-coupling was to represent the antenna with an equivalent pi network at the frequency in question and measure the impedance into Slot No. 1 under both open and short circuit conditions of Slot No. 2. In this way the coupling coefficient between Terminals 1 and 2 of the pi equivalent could be determined.

With the four-probe arrangement and an electrical phasing of 90° at center frequency, prototype antenna pattern measurements (patterns obtained with the antenna in a simulated, curved ground plane) indicate ellipticities of 1 db or less directly overhead, and maximum to minimum differences of radiated power in the horizontal plane to be less than 3.5 db over the band. The prototype configuration used to obtain these results was a crossed slot of the dimensions previously used for a single slot, and the improved probe. The radome used was of 5/16 inch moulded epoxy fiberglass. Each pair of probes were fed in parallel, with one pair phased 90 electrical degrees from the other by a quarter wave section of coaxial cable. The power was divided between slots by a broad band hybrid ring assembly designed and manufactured by Dynatronics, Inc. During the production of nine (9) units, the antenna VSWR in no case exceed 1.30:1 over the 215 to 260 Mc telemetry band. Figure 7 is a view of the prototype antenna without a radome.



Figure 7

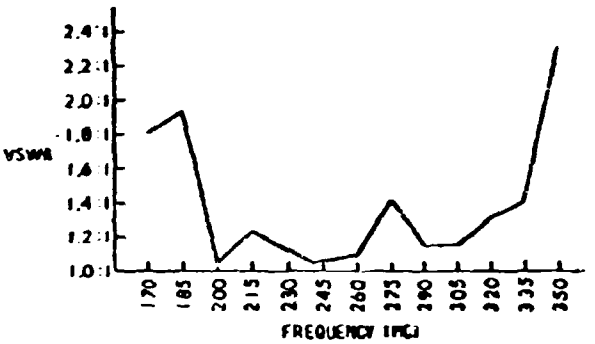
PROTOTYPE DESIGN EVALUATION TESTS

TEST DESCRIPTION

Using a simulated ground plane, the VSWR of the prototype and each production antenna were measured at 215, 240, and 260 Mc. The gain of the prototype antenna referred to a circularly-polarized source was determined by the dipole substitution method. Ellipticity measurements were made at 0°, 30°, 60° and 90° in elevation at each azimuth position of 0°, 45° and 90° on the prototype antenna.

GAIN MEASUREMENTS

The antenna gain was obtained by measuring the horizontal and vertical patterns of the antenna with a known power input. A standard dipole was then substituted for the prototype crossed-slot, and a pattern was run to determine its radiation characteristics. By a correlation of areas within each pattern, the gain of the antenna was readily determined. It was found that the antenna gain directly overhead was + 3.9 db and the gain at the horizon was -2 db; both gain figures referred to a circularly-polarized isotropic source.



VSWR VERSUS FREQUENCY

Figure 8.

VSWR MEASUREMENTS AND PATTERNS

The VSWR of the prototype antenna was measured and graphical results are shown in Figure 8. As can be seen, the antenna exhibits VSWR of less than 2:1 and should produce good radiation coverage from about 170 Mc to 350 Mc which is a frequency bandwidth of 2:1. The unit is, therefore, operable over a frequency range of four times that required for the present telemetry band of 215-260 Mc.

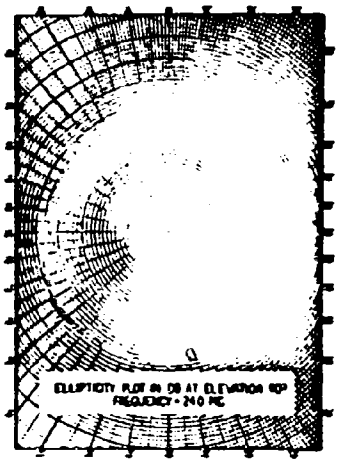
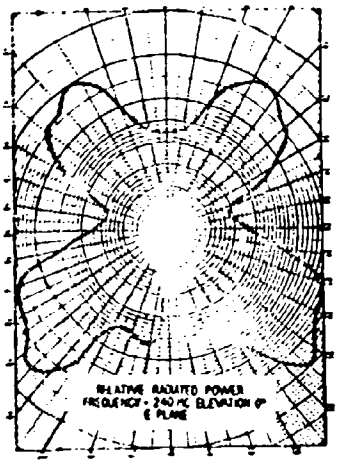


Figure 9.

Figure 9 is a typical prototype pattern taken directly overhead with the antenna installed in a simulated ground plane, indicating that the antenna has excellent axial ratio directly overhead. In many cases the axial ratio is good to about 40° or 50° from the vertical axis of the aircraft. The average circularity at 30° from the zenith is approximately 4 db and goes to roughly 15 db at the horizon.

SCALE MODEL PATTERNS

Scale model pattern measurements were made, utilizing a 1/24th scale C-54 aircraft. Since it was physically impossible to build a scale model antenna with all four probes driven, two probes were driven in a single slot while the other slot was not excited. Owing to the symmetry of each slot with respect to the aircraft, the actual scale model radiation patterns were easily obtained. Figure 10 is a typical



E field radiation pattern taken in the horizontal plane of the scale model.

TEST RESULTS AND CONCLUSIONS

The deepest antenna pattern hole referred to a circularly-polarized isotropic source occurs at a frequency of 260 Mc and is directly off the tail of the aircraft and has a depth of 12 db. At this null point, which is at the horizon, the reserve signal strength over a 500 mile range is calculated as follows:

Missile ERP/Antenna (20w)	43 dbm
Missile Antenna Gain (Half wave slot)	2 dbm
	45 dbm
Path Attenuation (500) miles	-138 db
Antenna Gain (at null point)	-12 db
Signal level at Pre-amp input (at worst possible conditions)	-105 dbm
Receiver Sensitivity (2 uv)	-101 dbm

It is seen that the signal strength is below the sensitivity of the receiver by about 4 db in the region very near the tail section of the aircraft and near the horizon.

However, null is extremely sharp. Within a few degrees on either side of the vertical stabilizer at this elevation, the signal level is above the sensitivity of the receiver. At this same point, and at a frequency of 240 Mc, the antenna signal is equal to that of the receiver sensitivity and at 215 Mc the antenna provides a 3 db margin above receiver sensitivity.

REFERENCES:

1. Antennas, John D. Kraus, McGraw-Hill Book Company, New York, 1950

2. Radio Engineering, Third Edition, F.E. Terman, McGraw-Hill Book Company, New York, 1947

3. Very High Frequency Techniques, Radio Research Laboratory, Harvard University McGraw-Hill Book Company, New York, 1947

4. Reference Data for Radio Engineers, Fourth Edition, International Telephone and Telegraph Corporation, New York, 1957

LIMITATIONS OF ANGULAR RADAR RESOLUTION

By: Ewald Eichler, U.S. Army Ordnance, Frankford Arsenal, Philadelphia 37, Pennsylvania

Angular radar resolution, its limits, and means for reaching or overcoming them are reviewed. The subject is linked to optical diffraction theory and to the resolution problem and the sampling theorem of communication.

In the absence of noise, resolution is limited to roughly the ratio d/λ (aperture over wavelength). This limit is reduced if the target dynamic range is greater than one. For average resolution $2d/\lambda$ is an exact limit. Uniform antenna illumination allows nearly optimum resolution. Noise reduces those limits, and target scintillation also poses a limit near d/λ. Different filters and radar techniques should be helpful if noise is the cause of diminished resolution. If one tries to exceed the limit d/λ one runs into forbidding expense. It is concluded that radar resolution beyond this limit is impractical. The application of time-domain filters and their realizability is also discussed.

I INTRODUCTION

The radar signal in response to a small point target is a replica of the squared antenna gain $g^2(\Theta)$. Because the same aperture is used twice, for transmission and for reception, the microwave amplitude pattern corresponds to an optical intensity diffraction pattern. RAYLEIGH defined resolution by reference to the separation between the center maximum and the first null of the diffraction pattern[1], and FREEDMAN defined radar resolution by reference to the beamwidth[2]. RAYLEIGH's criterion gives resolution Res = d/λ for a one-dimensional aperture and Res = $.82d/\lambda$ for a circular aperture of uniform illumination. FREEDMAN's criterion gives slightly smaller values.

Resolution or resolving power can be defined as the inverse of the minimum separation of two or more point objects which can still be seen apart with some instrument, such as a telescope or a radar system. Although this appears to be a sensible definition of an important property, it leads to some difficulties.

In communication resolution could similarly be defined by reference to a sequence of impulses which can still be separated or identified after passing through some system subject to distortion and noise. There is obviously no other way to resolve this signal than by noting that it consists of different parts which are well concentrated in time. A filter introduced for the enhancement of resolution has to bring out this property. Because each communication system has only a finite bandwidth an impulse is always being smeared out infinitely in time. Rigorous resolution is impossible therefore, and some arbitrary criterion must be introduced. The effects of finite bandwidth and noise will be looked into separately; finally a time-domain filter is investigated.

II. LIMITATIONS DUE TO FINITE BANDWIDTH

The signal $si2\pi Wt$ (si x being an abbreviation of $\sin x/x$) is the impulse response of an ideal low-pass of bandwidth W. The signal $si^2\pi Wt$, which for $Wt = d/\lambda \cdot \sin\Theta \doteq d\Theta/\lambda$ equals the squared antenna gain $g^2(\Theta)$ of a one-dimensional antenna of width d and uniform illumination[3], requires the same bandwidth W but a triangular spectrum instead of a flat one. RAYLEIGH's resolution of the first signal is twice that of the other one; FREEDMAN's criterion gives different values, too, although the two signals could be generated from one another (at least approximately) by a linear filter. The filter which would generate the first signal is usually called the "equalization" filter.

A time function and its spectrum (for instance an impulse response and the corresponding system function) are Fourier mates for which a general uncertainty relation $\Delta t \cdot \Delta f \geq 1$ holds[4]. The uncertainties Δt and Δf are (again somewhat arbitrarily) defined as $2\sqrt{\pi}$ times the mean normalized variance in the time and in the frequency domain, respectively. Thus defined, the uncertainty of $si^2\pi Wt$ is infinite, but the uncertainty of $si^2\pi Wt$ is $\sqrt{3/\pi}/W$ and its uncertainty product is 1.1, very close to the limit. Within a limited frequency band the smallest uncertainty product is obtained for a "half cosine" band shape[4]. It appears, hence, that uniform radar antenna illumination is not far from optimum. A filter with the purpose of increasing resolution should aim at creating a "half cosine" shaped rather than an "equalized" frequency band.

From all this it is clear that resolution must be proportional to bandwidth. There may be some

[1] F.A. JENKINS and H.E. WHITE: Fundamentals of Optics. McGraw-Hill, N. Y., 1950

[2] J. FREEDMAN: Resolution in Radar Systems. P.I.R.E. 39, 813-18 (1951)

[3] S. SILVER, ed. Microwave Antenna Theory and Design. Rad. Lab. Ser. 12, Mc-Graw-Hill, N.Y., 1949

[4] D. GABOR: A Summary of Communication Theory. W. JACKSON, ed. Communication Theory. p 1-23, Butterworth, London, 1953

dispute about the proportionality factor, but a good choice is to write Res = W. More accurately, this appears to be the limit or the obtainable resolution with a system of bandwidth W.

If there are several objects with an optical system their intensity patterns will add linearly, because of incoherence. If there are several radar targets $h(\vartheta)$ their signals will also add linearly,

$$e(\theta) \sim \int_{-\infty}^{\infty} g^2(\vartheta - \theta)\, h(\vartheta)\, d\vartheta$$

($\sim$ denoting proportionality). After taking the Fourier transform, $E(\phi) \sim \Gamma(\phi)H(\phi)$ follows, (where E and H are the transforms of e and h, respectively). $\Gamma(\phi)$, being the transform of the squared gain is

$$\Gamma(\phi) \sim \int_{-\infty}^{\infty} G(\psi)G(\phi - \psi)\, d\psi$$

or proportional to the convolution of the gain transform $G(\phi)$. The gain itself is[3]

$$g(\theta) \sim \int u(2\pi x/\lambda) \exp(j2\pi x/\lambda \cdot \sin\theta)\, dx.$$

Taking the Fourier transform

$$u(2\pi x/\lambda) \sim \int g(\theta) \exp(-j2\pi x/\lambda \cdot \sin\theta)\, d\sin\theta$$

is obtained, and this is approximately equal to

$$\int_{-\infty}^{\infty} g(\theta) \exp(-j2\pi x\theta/\lambda)\, d\theta \sim G(x/\lambda)$$

because the gain is very small where $\sin\theta$ deviates appreciably from θ. Hence, the transform of the gain is proportional to the illumination, and the radar system function $\Gamma(\phi)$ proportional to its convolution, if we set the "frequency" $\phi = x/\lambda$. In particular, because the illumination goes to zero for $|x| > d/2$, the radar represents a low-pass filter of "bandwidth" d/λ and hence of possible resolution d/λ.

More accurately, the radar represents two such filters, one for the cosine and one for the sine microwave component (which can be separated by a phase-sensitive detector). Take two point targets of different phase: the microwave amplitude will then be anywhere between the sum and the difference of the individual returns, depending on the relative phase. Unless the relative phase is known a priori, nothing is gained, and the two targets cannot be resolved better than if they had the same phase.

So far, signals of different amplitude have not been mentioned. It is clear that a pulse of larger amplitude will spread out further in time although its uncertainty stays the same. The pulse $A\ si^2\pi Wt$ takes $\sqrt{A}$ times longer than $si^2\pi Wt$ to decay to the same level. Taking this signal as representative, it is found that the resolution formula should be modified to $Res = W \cdot \sqrt{D}$ where D is the largest amplitude ratio which occurs, or, the dynamic range.

One must be careful not to confuse resolution with accuracy. It seems possible, by accurate inspection of the signal generated by two impulses, to recognize separations much smaller than 1/W.[5] The composite pattern might be looked up in a catalog of all possible patterns. To determine the separation between two pulses is a task of accuracy, however, and can be driven well beyond the resolution limit. First it must be known, that the pattern was generated by just two impulses, whether this knowledge is obtained by the process of resolution or given. Even the catalog, apart from the impossibility to compile infinitely many patterns, can only serve accuracy or tell "there is not only one target; there are not only two targets, etc".

It has been shown, that no linear filter can improve resolution beyond a value around $W/\sqrt{D}$. Can a non-linear filter help? A squaring device, for instance, reduces the uncertainty of a pulse like $\exp(-t^2)$ by $\sqrt{2}$ and increases the dynamic range from D to D^2. Hence resolution is increased by $\sqrt{2/D}$ or not at all if $D > 2$. A greater receiver dynamic range is required. A squaring device accentuates the dip between pulses; the minimum separation between two pulses $si^2\pi Wt$ which makes for a dip is .7/W, and the minimum spacing between two maxima for signals of bandwidth W is, of course, 1/2W. A threshold device, another non-linear filter, admits only a single amplitude. Little resolution improvement is thus expected from non-linear filters and it is also expected that they will be easily disturbed by noise.

The sampling theorem states that only one amplitude per time interval 1/2W can be specified in the average[6]. It follows that average resolution must be limited as $Res_{av} \geq 2W$. Even the mentioned catalog will be useless beyond this limit. In many cases the original signal is not a sequence of impulses but a continuous function. Then the concept of resolution becomes useless or has to be replaced by a concept of fidelity. If fidelity is defined by reference to the least mean square deviation, the "equalized" signal is the best possible approximation to the original. If the original signal contains only frequencies up to W, no detail is lost, of course.

---

[5] D. GABOR: Collecting Information on Partially known Objects. Z. KOPAL, ed. Symposium on Astronomical Optics, North Holland, Amsterdam, 1958.

[6] S. GOLDMAN: Information Theory, Prentice Hall, N. Y., 1953.

III. LIMITATIONS DUE TO NOISE

In a radar system at least the following kinds of noise can limit or reduce resolution: (1) target scintillation, (2) clutter, (3) atmospheric turbulence, (4) antenna instability, (5) frequency instability, and (6) receiver noise. From an inspection of the $\sin^2 \pi Wt$ pulse one can expect noise to reduce resolution to very roughly $W\sqrt{1 - 2A_N/A_S}$, where A_N is the noise amplitude and A_S is the signal amplitude.

All noise components beyond the signal bandwidth $W = \dot{\theta}d/2\pi\lambda$ ($\dot{\theta}$ = antenna scan rate) can, of course, be eliminated by a filter without loss in resolution. (Those due to target scintillation may, however, carry significant information and can be utilized in aural presentation.)

Target scintillations are fluctuations in the microwave amplitude and phase and are caused by target motion. A scintillating target, such as a propeller aircraft, can cause a radar signal with several distinct peaks which can be mistaken for separate targets. Target amplitude scintillations can be eliminated by monopulse radar[7]. With two targets of fluctuating phase present, the radar signal amplitude fluctuates between the sum and the difference of the individual signal amplitudes. This fluctuation could be reduced by integration over several antenna sweeps. But this is necessary only if one tries to surpass the d/ limit, because for greater target separation there is essentially only one target in the beam at one time.

Clutter is caused by radar reflections from other objects near the proper target or targets. It can be reduced only by use of some differentiating property. Polarization techniques could help if target and clutter have different reflectivities for different polarization, Doppler techniques if they are moving differently, and also monopulse[7]. Atmospheric turbulence causes the beam to bend and to spread and there appears no simple remedy[8]. Antenna instability should have a similar effect. Frequency instability would disturb a Doppler scheme.

Gaussian noise can best be suppressed by the "matched" filter which maximizes the ratio of instantaneous signal power to average noise power[6]. For noise flat with frequency it is almost the inverse of the equalization filter. The WIENER filter restores a continuous signal of fixed statistical properties in Gaussian noise according to the least mean square condition[6]. Noise within the signal band is suppressed by integration over several antenna sweeps provided the targets are sufficiently stationary. This can be done by means of comb filters which are tuned to the scan frequency $\dot{\theta}/2\pi$ and its harmonics.

IV. A TIME-DOMAIN FILTER

A time-domain or delay-line filter is a linear filter using a delay line and one or two summing amplifiers (fig.). The line is tapped at regular intervals T, the a_i and b_i are real multipliers. The response function is

$$\frac{E_0}{E_1} = \frac{a_0 z^N + a_1 z^{N-1} + \ldots + a_N}{z^N - b_1 z^{N-1} - \ldots - b_N}$$

where $z = \exp(Ts)$ has replaced the complex frequency s.

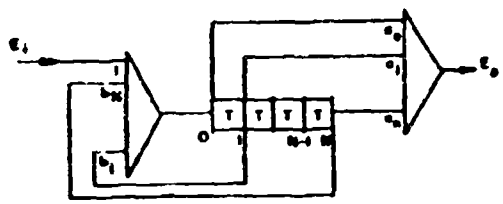


Figure. Time-Domain Filter

It seems that any desired function could be obtained as the output of such a filter by choosing the coefficients such that the denominator becomes the z-transform of the given input and the numerator the z-transform of the desired output. But for the filter to be stable, all poles must lie within the unit circle. They are zeros of the denominator polynomial and their location with respect to the unit circle is found by means of the SCHUR-COHN criterion[9].

Such a filter had been proposed to increase radar resolution, by choosing as numerator and denominator respectively the (sampled) point target response of a high and a low resolution radar. Because the gain is an even function, the denominator must have the shape $D(z) = d_0 + d_1 z + \ldots + d_1 z^{N-1} + d_0 z^N = z^N D(1/z)$. Hence, with z_i a

---

[7] S. Y. CHANG: An Experimental Study of monopulse Radar for Ground-clutter Discrimination. 1960 I. R. E. Mil. Electronics Conference Record.

[8] C. M. ANGULO and J. P. RUINA: Antenna resolution is limited by Atmospheric Turbulence. Presented at URSI, May, 1957.

[9] M. MARDEN: The Geometry of the Zeros of a Polynomial in a Complex Variable. Am. Math. Soc., N.Y., 1949

zero $1/z_1$ must be a zero, too, and as many zeros must lie outside the unit circle as there lie inside. There must be N zeros altogether; even $|z_1| = 1$ is not safe for stability, and the filter is unrealizable.

In the frequency domain the unrealizability of the filter is also evident. The spectrum of the high resolution radar signal extends to proportionally higher frequencies, and these frequencies cannot be generated by a linear filter. In a non-feedback type filter all the b_i are zero. If the first part of the output function is specified to be a time-compressed radar signal, a tail results which is several times higher than the desired part of the output.

IV. CONCLUSIONS

From the Fourier uncertainty relation of communication it is concluded that the bandwidth of a linear system and its optimum resolution are not only proportional but about equal. Antenna theory shows that a radar represents a low-pass filter. The possible resolution of radar signals and of incoherent optical signals is both d/λ, the ratio of aperture width over wavelength. These relations are valid for the Fraunhofer region and large d/λ. Whether focusing might improve resolution is not discussed. One difference between optics and radar is that the derived relations hold for the two quadrature components of the radar signal rather than for its amplitude. A filter which presupposes linearity should deal with the two components separately. This requires a phase-sensitive detector. But this appears not critical if the target spacing is beyond λ/d.

Accuracy and resolution should not be confused; resolution is a prerequisite of accuracy. If the object does not consist of isolated points as in astronomy or in a sky-looking radar, but is rather a continuous function as may be the case in a ground-looking radar, resolution does not make sense and is better replaced by a fidelity requirement. The radar of maximum resolution is not also the radar of least mean square error.

Resolution is reduced by $1/\sqrt{D}$ if the signal dynamic range D is greater than one. Noise diminishes resolution by another factor. The usefulness of non-linear filters for resolution enhancement appears quite small and is related to dynamic range and bandwidth. Different kinds of noise which degrade resolution below d/λ can be met with various filter and radar techniques. A certain time-domain filter which had been proposed for resolution enhancement is shown to be unrealizable.

SPACE SIMULATION WITH HIGH GAS RELEASE RATES

By: Mr. W. W. Balwanz and Mr. J. M. Singer, Radar Division, U. S. Naval Research Laboratory

INTRODUCTION

No satisfactory instrumentation for simulating the pressures of high altitudes with high gas-release rates is available on the surface of the earth today. The scientist of the future, working on space platforms, may be satisfied with the pumping capacity of outer space, but until the time of ready availability of such facilities, we here on earth must provide our own evacuated work areas.

Facilities now in existence for maintaining low pressures when operating rocket motors and jets under simulated high-altitude conditions are limited to pressures of one millimeter of mercury or greater, i.e., simulated altitudes of 30 miles or less. The cost of extending the pumping capacity of such facilities to simulate higher altitudes using conventional pumping systems is exorbitant. The cost of conventional pumps for altitude simulation with a 10-pound-thrust rocket motor is illustrated in Table 1. Costs for larger gas-release rates would be proportionately higher. If the conventional pumping methods were the only ones available, it would probably be more economical to go to outer space for low-pressure studies requiring high gas-release rates.

Table 2
Pumping Methods

1. Mechanical displacement
   a. Rotary pumps
   b. Blowers
2. Momentum transfer
   a. Diffusion pumps
   b. Jet ejectors
3. Valence bond attachment
   a. Chemical reactions
   b. Crystalization
   c. Condensation
   d. Adsorption
   e. Absorption
4. Mechanical bond attachment
   a. Adsorption
   b. Absorption
5. Matrix trapping
   a. Interment
   b. Absorption

Table 1
Conventional Pump Requirements for a 10-Pound-Thrust Rocket Motor Operating at Simulated Altitude

Pressure (mm Hg)	Altitude (Miles)	Volumetric Pumping Rate		Number of Conventional Pumps Required		Pump Cost (Dollars)	Electric Energy (kw)
		cfm	l/sec	3-stage ejector diffusion pumps	48-inch diffusion pumps		
10^{-4}	68	2.6×10^{5}	1.2×10^{5}	62	28,000	2.70×10^{8}	3.74×10^{8}
10^{-6}	101	2.6×10^{7}	1.2×10^{7}	62	478,000	1.59×10^{9}	7.36×10^{7}

Improved pumping methods are urgently needed, since the space platform will probably not be available to the average scientist within the next several years.

PUMPING METHODS

Studies to explore and evaluate the efficacy of possible improved pumping techniques have been initiated at NRL. Early indications are that material gains in reducing the cost of high-pumping-rate systems at high simulated altitudes will be achieved. Pumping methods may be classified as shown in Table 2.

The development of mechanical-displacement pumps is highly advanced, and appreciable gains in the efficiency of such pumps are not likely. A similar conclusion applies to the second of these methods, the momentum-transfer pumps. Recent developments in the efficiency of diffusion pumps give performance approaching the theoretical maximum of such devices. Further applied studies are needed to determine the extent of utilization of the momentum of rocket exhaust gases in the lower pressure regions of interest, but material gains in efficiency over that attained in the higher pressure regions are not anticipated.

Efficient utilization of the remaining methods, however, give promises of major increases in pumping efficiency. Reduction in the costs of high-pumping-rate systems by orders of magnitude appear to be feasible. It appears likely that both the initial and the final costs may be reduced to a fraction of a percent of costs using the conventional systems. Although knowledge presently available is inadequate for a full assessment of the gains to be achieved, experimental studies in some areas have shown reductions in cost as large as three orders of magnitude, and theoretical studies indicate greater gains possible.

The temperatures required to provide crystallization and condensation for various gases at reduced pressures are shown in Fig. 1. Liquid nitrogen temperatures at atmospheric pressures are suitable for removing water vapor, carbon dioxide, and less volatile substances to pressures of 10^{-7} or less. Liquid hydrogen temperatures are required for removing oxygen, nitrogen, and carbon monoxide to the same low-pressure region. Liquid helium temperatures are required to remove the last major constituent of rocket exhausts, hydrogen, at this low pressure. These curves show the limiting pressures attainable for various pure gases, but they do not show the pumping rates achievable in practical systems where gas samples usually contain appreciable amounts of impurities. Such impurities are the greatest source of uncertainty in the interpretation of the results of experimental measurements. For example, liquid nitrogen temperatures on pure carbon dioxide would provide a pressure of less than 10^{-12} atmospheres. A noncondensable mixed with the carbon dioxide in the extremely small ratio of one part in ten million would result in an ultimate pressure five orders of magnitude larger.

Chemical reactions induced to provide a molecule with lower vapor pressure have been shown to be effective in a number of instances. For example, in some pressure regions, the excess hydrogen in a rocket exhaust may be burned with externally supplied oxygen to provide the lower vapor pressure water molecule. Although this is not the principal mechanism, some chemical reaction probably occurs in ion pumps and in the gettering process used in vacuum-tube evacuation.

Both adsorption and absorption pumping methods have been demonstrated to be effective in some regions of interest. However, the pumping rates are very low at low pressures, and the techniques do not appear to be attractive for use in high-pumping-rate systems. The adsorption pumping rate can be materially increased by the use of low-temperature adsorbents, and in some cases this method may be used to supplement other techniques. Appreciable interment of noncondensables occurs when freezing out large quantities of gas from the exhaust of a rocket motor or other gas source. These interred molecules are trapped in the crystal matrix. In many cases, such interment is only temporary, the molecules eventually finding a way out unless some bond is established within the matrix.

Although appreciable knowledge is available on most of these methods (1-4), further experimental knowledge is needed to permit the efficient design of a large high-pumping-rate system. Exploratory experimental studies have been conducted to evaluate the potentiality of some of these methods, and further studies are continuing.

EXPERIMENTAL RESULTS

A cold-wall tank about ten feet long and two feet in diameter, identified as a two-phase pump, has been used for experimental studies. The tank, covered by an insulating jacket, is illustrated in Fig. 2. Liquid nitrogen sprinkled over the tank provides the cooling. Initial evacuation is achieved using an MCF-300 diffusion pump and a small forepump.

Calrod heating units wrapped around the tank are used to assist in outgassing. After pumping down to a pressure of 10^{-6} mm Hg or less, the tank is cooled to liquid-nitrogen temperature and test gases are admitted at controlled rates. Conventional flow meters are used for measuring the flow rates. The pressures attained as a function of flow rate for various gases is shown in Fig. 3. The pumping rate (for air) of a commercially available 48-in., 3-stage, fractionating diffusion pump is shown on the same plot for the purpose of comparison. The water vapor pumping rate of the two-phase pump, as shown by the point at the upper left on this figure, is greater than that of the diffusion pump by more than three orders of magnitude in the region investigated. Comparable and higher pumping rates may be possible for all gases, provided that colder surfaces are used.

At these large flow rates, a large number of noncondensable gas molecules are interred by the continued avalanche of crystallizing particles. Two different gas inlets into the system, one for noncondensable gases and the other for condensable gases, were used to measure the effect. The pressure-reduction ratio achieved by this mechanism varies from negligible to about 16 to 1, depending upon the gas combination. Table 3 shows the results of some of these studies. Further studies are required to establish the variation of the pressure reduction ratio with temperature.

APPLICATION

Applications for high-pumping-rate low-pressure systems are found throughout the fields of scientific endeavor. Although knowledge is still inadequate to permit the design of efficient engineering systems, an idealized example may be used to illustrate the potential of improved pumping methods. For example, consider the operation of the rocket motor of Table 1 at a simulated altitude of 100 miles. A specific impulse of 240 pound-seconds gives a flow rate of 2.4 pounds of exhaust gases per minute of operation. Assume that these gases pass through three successive chambers, as illustrated in Fig. 4, the first cooled with liquid nitrogen, the second with liquid hydrogen, and the third with liquid helium. The gases will condense or crystallize in the cooled chamber, carbon dioxide and water vapor in the first, carbon monoxide and oxygen in the second, and hydrogen in the third, down to a pressure of 10^{-6} mm Hg or less. All gases in the rocket exhaust except helium are condensed or crystallized in these chambers. Any small helium concentration which might be present can be evacuated with a relatively small conventional pumping system.

Less than 50 liters of liquid nitrogen, one liter of liquid hydrogen, and 2 liters of liquid helium would be used per minute of rocket-motor operation, assuming interment trapping to be 90-percent efficient. The cost of the liquefied gases is about 1000 times less than the

Table 3

Capture of Noncondensable Gases with Condensables

Noncondensable Gas	Condensable Gas		Chamber Pressure (mmHg)		Pressure Reduction Ratio
	Type	Flow rate cfm at STP	Noncondensable Gas only	Both gases	
H_2	NH_4	0.80	6.2×10^{-4}	7×10^{-5}	8.8
H_2	NH_4	1.05	5×10^{-4}	3×10^{-4}	16.6
He	NH_4	---	No reduction		0.0
H_2	SO_2	---	No reduction		0.0
H_2	CO_2	0.9	4×10^{-4}	2.5×10^{-4}	1.6
Air	SO_2	0.6	2×10^{-4}	6.0×10^{-5}	3.3
Air	SO_2	0.5	1.6×10^{-4}	7.0×10^{-5}	2.3
Air	Freon 12	0.17	5×10^{-4}	1.8×10^{-4}	2.8
Air	Isobutane	0.025	1.6×10^{-4}	9×10^{-5}	1.8

cost of the electricity required for operating the pumps for the conventional system as illustrated in Table 1. Estimated costs of the initial pump installation show similar savings.

CONCLUSION

It is concluded that improved pumping techniques orders of magnitude more efficient than those conventionally used can be constructed. Utilization of these improved techniques give promises of better space simulation than presently attained with high gas-release rates in the space facility. Further studies are needed to make efficient design a reality.

1. Dushman, S., "Vacuum Techniques," New York:Wiley, 1949
2. Scott, R.B., "Cryogenic Engineering," New York:Van Nostrand, 1959
3. "Handbook of Chemistry & Physics," Cleveland:Chemical Rubber Publishing Co., 1955
4. Forsythe, W.E., "Smithsonian Physical Tables," Washington Smithsonian Institute, 1954

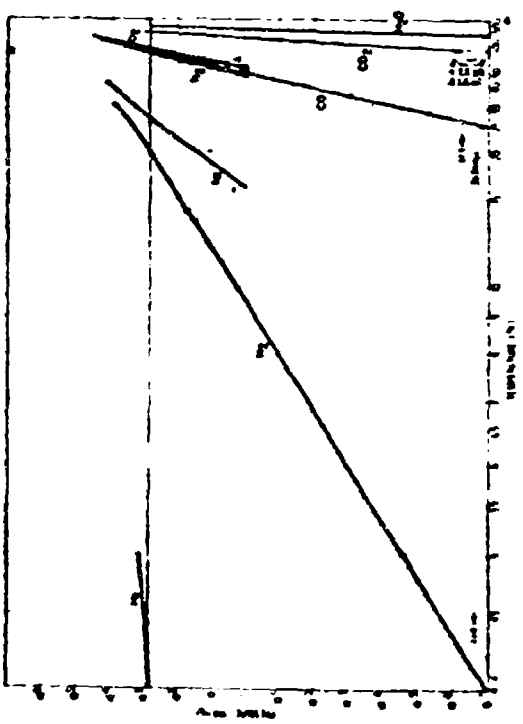
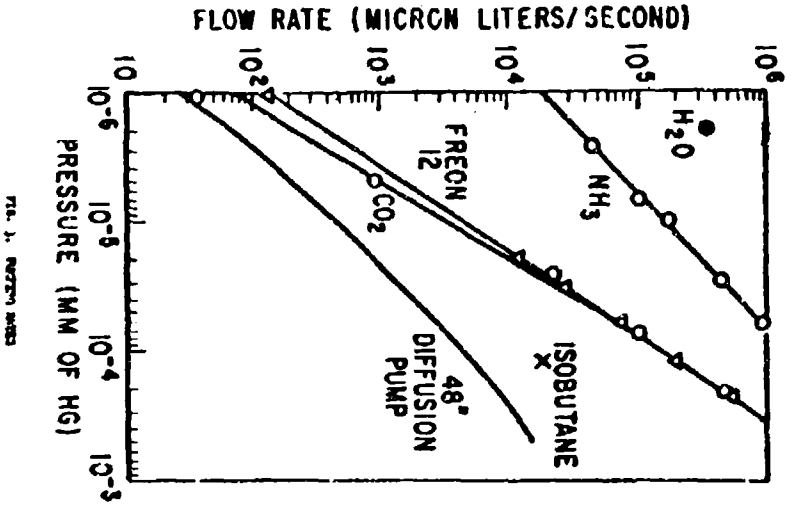


Figure 1



FIG. 2. AN EXPLORATORY TWO-PHASE PUMP



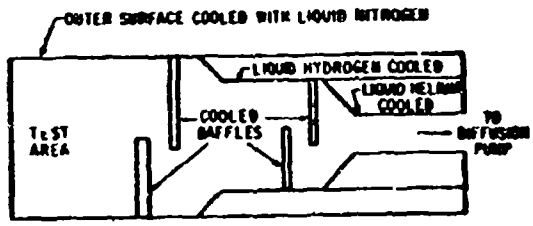


Figure 4 - Schematic of a Proposed Two Phase Pump

TANK VS TANK SYNTHETIC GUNNERY TRAINER

By: Irvin Friedland, Project Engineer, U. S. Naval Training Device Center

Introduction

This paper covers the development program performed under Contract No. 61339-212 for the U. S. Naval Training Device Center, Port Washington, New York by the Armour Research Foundation at Chicago, Illinois. The objective of the program was the development of a scoring system for use in training exercises which involve tank to tank gunfire exchange. In particular, what was desired was a simple device (scoring system) which could be easily associated with actual tanks so that upon the "firing" of a simulated round, the scoring system could determine whether or not an actual round would have scored against the target tank and would automatically indicate the success of such a round by the ignition of a smoke bomb on the target tank.

General System Description

The system is best introduced by considering two major subsystems, each of which can be discussed independently. These two subsystems are defined according to function and are concerned respectively with the problems of hit determination and hit communication. By hit determination is meant the process of deciding whether or not a particular simulated round would have resulted in a hit on the tank which is the object of fire. Because of certain characteristics of the system, this decision is effected by means of equipment on the tank which fired the round (referred to herein as the attacking tank, as opposed to the target tank which is the object of fire). The hit communication subsystem is then concerned with communicating the information (from the attacking to the target tank) that a hit has been scored.

The hit determination subsystem operates as follows. At the instant the gun trigger is pulled an extremely high energy, short duration light pulse is initiated by means of a flash tube. This flash tube is positioned on the axis of a narrow angle optical projection system which is mounted rigidly on the gun barrel and which has been boresighted with the gunner's sight. The superelevation cam in the superelevation computer is removed so that if the gun is correctly aimed, the axis of the projection system will intersect some portion of the target tank and the projected light pulse will strike the target. The target tank turret is essentially covered with a retro-reflective material known as Scotchlite. This material has the property that it will reflect light back along the line of incidence rather than in the usual mirror fashion. By virtue of this retro-reflective covering, the system has the property that if any portion of the target tank is intercepted by the axis of the projection system, light will be reflected along this line back to the attacking tank. A second optical system, boresighted with the first and containing a suitable photodetector will provide an output pulse if the gun is suitably aimed. Failure to receive such a pulse is in fact an indication that the gun is improperly aimed.

Although it turns out that in practice a system employing only Scotchlite as a retro-reflector is noise limited to ranges less than the desired 2500 yds, this problem is easily overcome by means of a set of corner cube retro-reflectors called cataphotes. The cataphotes reflect a high signal level at the maximum range. The reason that the cataphotes will not serve by themselves is that at the shorter ranges (under about 1000 yds) it is possible that the optical axis of the receiving system will intersect some point on the target tank sufficiently far from the corner cubes that they are excluded from the field of view. At these ranges, however, the return from the Scotchlite is adequate. Figure 1 shows the relative signal strength of the Scotchlite as a function of angle and figure 2 shows the signal strength as a function of the range.

In order to determine whether or not a hit would have been scored by the simulated round it is essential to know not only whether or not the gun is correctly aimed, but whether or not the tank commander's range estimate is accurate. In the device to be described, the true range is compared with the estimated range to determine if the latter is sufficiently close to correct for a hit to have resulted if a round were fired. When this condition is met, the hit determination system "decides" that a hit would have been scored.

Due to the finite extent of the target, a small error in the range estimation will still result in a hit but the impact point will be higher or lower than the aim point. Thus for each estimated range there is a band of true ranges over which a hit will occur. In general, there are actually two such range bands or hit zones - Zone #1 which extends out from the gun and Zone #2 which brackets the estimated range. For estimated ranges less than some "critical" value, the two hit zones merge to form a single hit zone extending from the gun out to beyond the estimated range.

The boundaries of the hit zones are a function of vertical target size (or permissible vertical error in target plane), estimated range, and projectile trajectory. The relationship between these factors is illustrated in Fig. 3, for the assumed condition of the gun muzzle and the target center being at the same elevation. Using this assumption and specifying a vertical miss distance and projectile type, the hit zone boundaries were calculated for the HVAP-T, M304 90mm cartridge and are shown as impact point loci in Fig. 4 for miss distances of both two feet and five feet. Data for the calculations were taken from Department of the Army Firing Table FT90-M-1, dated June 1957. The hit zone boundaries can be accurately represented by the solid straight line segments shown in Fig. 5. The abrupt step in the upper limit (Range Function #2) of the hit zone shown in this figure results from the fact that for estimated ranges less than 500 yards no superelevation is applied to the gun.

The operation of the hit communication subsystem is as follows. A continuously modulated, infrared, omnidirectional beacon is mounted on top of the target tank. The beacon modulation frequency is different for each of the tanks participating in the exercise. Infrared radiation is employed so that the beacon will not reveal the tank location.

A second optical receiving system (in this case one which employs an infrared detector) is located on the attacking tank and aligned with the axis of the hit determination subsystem previously discussed. As soon as the gun is trained on the target tank, radiation from the modulated beacon enters the infrared receiving system and an electrical signal is generated. If the hit determination system "decides" that a hit would have been scored the output of the infrared receiver is connected to the input terminals of a radio frequency transmitter. This radio frequency transmitter is capable of communicating with all participating tanks. In the event of such a hit the radio frequency transmitter is keyed on for approximately 0.15 sec during which time the message it transmits is simply the output of the infrared photodetector. This signal is of course received by all of the participating tanks. Only one of the tanks however identifies the received signal as its own frequency, i.e., the one from which the signal is in fact being derived. The fact that a reference signal is available at each tank for use in decoding or identifying signals of its own frequency makes the use of a phase sensitive demodulator possible in this application. The reference phase for the phase sensitive demodulator is derived directly from the beacon on top of the tank. The input to the phase sensitive demodulator is the demodulated output of the radio frequency receiver. If the input to the phase sensitive demodulator and the reference signal coincide in phase and frequency, the phase sensitive demodulator will have an output which is taken as an indication of a hit.

It is interesting to note that if the scoring situation were one in which target range is not important, then the system described herein as the hit communication system could be employed as the entire scoring system.

Detailed System Description

The optical system consists of five major components: 1) a projector consisting of a flash tube and projector optical system; 2) a receiving optical system consisting of collecting optics and photomultiplier detector; 3) a communication receiving system consisting of collecting optics and a lead sulfide cell as sensor; 4) retrodirective reflectors on target tank consisting of Scotchlite and corner cube reflectors and 5) an omnidirectional modulated light source on target tank.

Since items 1, 2, and 3 must all be aligned with the axis of the gun barrel all of these items were included in a single assembly which was mounted on the gun barrel, Fig. 6. Items 4 and 5, on the other hand, must be visible from almost any aspect and were therefore combined in a single assembly which was mounted on top of the cupola, Fig. 7. Fig. 8 shows the location of these assemblies on the tank and provides an overall view of the employment of the device as a simulated combat situation.

In order that accurate range information may be obtained from the optical radar, the rise time of the light pulse must be very short (0.1 microsecond). The source must be of high intrinsic brightness since the flux incident on the phototube is a function of the brightness of the source. Both conditions of short duration and high brightness were adequately satisfied in GE flash tube FT 230, operated with 0.1 microfarad at just under breakdown voltage (2700-3200 V).

Since the detector used in the optical radar must receive pulses of radiation with short rise times (0.1 microsecond) and maintain the waveform of the signal, a photoemissive device must be used. A multiplier type phototube DuMont 6467 was chosen for convenience since it provided a high signal output without high gain amplifiers.

In order that the optical communication system may operate with visual security, either infrared or ultraviolet radiation must be used. Infrared was the choice because of the high scattering of ultraviolet as well as the difficulty of making an ultraviolet source invisible to young observers. A lead sulfide cell was chosen as the detector, and the transmission of the filter over the light source chosen so as to permit operation from 0.7 to 2.8 microns.

The range functions No. 1 and No. 2 shown in the hit determination block diagram of Fig. 9 have voltage values which are determined by the range set into the T-31 ballistic computer by the Tank Commander. The range function No. 1 represents the lower boundary of the hit zone drawn in Fig. 5, with 50 millivolts corresponding to 1 yard on the ordinate scale. The range function No. 2 represents the upper boundary of the hit zone. Since the input shaft of the T-31 computer is logarithmic and rotates 23.5 times for a 20 db change in range (5000 to 500 yards), the rotation is transmitted to a 20 db logarithmic potentiometer by a chain and series of gears which have a 23.5 to 1 total ratio. A linear relationship between the voltage at the output of the potentiometer and the range set on the T-31 computer results. The sloped portions of the range functions No. 1 and No. 2 are then obtained by using suitable dividing networks at the output of the potentiometer. The discontinuities that the range functions No. 1 and No. 2 contain are produced by attaching a cam to the potentiometer shaft and switching the input to the comparators No. 1 and No. 2 from the potentiometer to appropriate dividing networks at the proper ranges.

In order to make 50 millivolts correspond to 1 yard, the ramp generator voltage must be at 50 volts when the light from the flash lamp has traveled 2000 yards (1000 yards to the target and 1000 yards return). The light will travel 2000 yards in $\frac{2000}{328 \times 10^6} = 6.1 \times 10^{-6}$ seconds, since light travels 328×10^6 yards per second. So the ramp voltage must rise at the rate of $\frac{50}{6.1 \times 10^{-6}} = 8.2 \times 10^6 \frac{\text{volts}}{\text{second}}$.

To demonstrate the operation of the hit determination system, it will be assumed that the tank commander estimated the range of a target tank at 1000 yards and that the gun is properly aimed at this target tank. When the gun is fired, the FT-230 gun discharge lamp sends out a high frequency pulse of light. The discharge of the lamp switches bistable multivibrator (B.M.V.) #1 to the position which we call "on". (The switching of the main elements of the hit determination system are illustrated on a timing chart in Fig. 10). The cathode follower of course is also switched to an "on" position. This in turn starts the ramp generator which builds up a voltage linearly with time at the rate computed above (8.2×10^6 volts/seconds). Furthermore, the inputs to And Gate #3 are now B.M.V. #1 "on" and B.M.V. #2 "off". This combination of inputs makes And Gate No. 3 receptive to a returning light pulse which is relayed to the gate from the photo-multiplier and its associated amplifier.

Since the target range estimated by the tank commander is 1000 yards, the range function No. 1 will have a voltage setting which corresponds to 740 (see Fig. No. 5) yards or $740 \times 50 \times 10^{-3} = 37$ volts. Then after $\frac{37}{8.2 \times 10^6} = 4.5$ microseconds the comparator #1 will produce a pulse. This pulse turns on bistable multivibrator (B.M.V.) #2. When B.M.V. #2 goes on, And Gate #1 becomes receptive to a returning light pulse, while And Gate #3 is no longer receptive. Range Function #2 will have the voltage setting which corresponds to 1170 yards (see Fig. No. 5) or $1170 \times 50 \times 10^{-3} = 58.5$ volts. Then after $\frac{58.5}{8.2 \times 10^6} = 7.1$ microseconds the comparator No. 2 will produce a pulse. This pulse turns off B.M.V. #1 and its associated cathode follower.

When the cathode follower is turned off, it turns off B.M.V. #2 which makes And Gate #1 unreceptive to a light pulse, and at the same time it turns on monostable multivibrator (M.M.V) #4 which makes And Gate #4 receptive to a returning light pulse. M.M.V. #4 remains on for 10 microseconds and then goes to the off position which makes And Gate #4 unreceptive to a light pulse. It should be noted that of the three And Gates, one and only one is receptive to the returninglight pulse at any particular instant of time throughout the critical 15 microsecond period following the firing of the simulator. If And Gate #3 is in a receptive state when the reflected light pulse from the target is received, the Gate puts out a pulse which triggers M.M.V. #5 which turns on the "over" light for approximately two seconds. If And Gate #1 is in a receptive state when the pulse is received, the Gate puts out a pulse which triggers M.M.V. #7 and M.M.V. #3. M. M.V. #7 turns on the "On Target" light for approximately two seconds. M.M.V. #3 turns on the transmitter for approximately 250 milliseconds. If And Gate #4 is in a receptive state when the pulse is received, the Gate puts out a pulse which triggers M.M.V. #6 which turns on the "short" light for approximately two seconds.

In the case of a "hit", the output produced by M.M.V. #3 when it is triggered turns on a radio transmitter for about 150 milliseconds. Modulation for the radio transmitter is provided by the output of the lead sulfide detector which senses modulated infrared radiation from the tank being fired upon. Thus the frequency of the modulation applied to the radio transmitter is precisely the same as that of the modulation of the infrared radiation. All tanks are emitting modulated infrared radiation continuously but each tank uses a different modulation frequency in the band of 300 to 1000 cycles per second.

Hit Receiver and Indicator

The radio receivers in all tanks are on at all times except for the receiver in a "firing" tank which is off for the 150 milliseconds that its transmitter is on. Thus all tanks, except the one "firing", receive any and all radio transmissions from other tanks.

The output of the radio receivers is a signal at the modulation frequency of the radio transmission. As shown in Fig 11, this signal is applied to a Phase Sensitive Detector (PSD) along with a reference signal provided by a photocell which views the tank's own modulated infrared source.

The action of the PSD can be considered as that of an instantaneous multiplier. Thus its instantaneous output is proportional to the instantaneous product of the two input signals. With two square wave inputs of different fundamental frequencies, the PSD output will consist of an ac voltage whose frequency components are sums and differences of the odd integer multiples of the two input fundamental frequencies and will have no dc component. However, when the two inputs are of precisely the same frequency, and in phase, the PSD output will consist of a dc level. If the phases of the two signals are not the same, the dc level will be lower (it becomes zero at 90 degrees and 270 degrees) and in addition, there will be an ac voltage whose frequency components are integer multiples of twice the fundamental frequency.

In the tank which has been fired upon, the two signals applied to the PSD will have precisely the same frequency since both signals are derived from the same modulated infrared source (via a radio link in one case). This is illustrated in Fig. 12. In all other tanks, the two signals will have frequencies that are different by at least 5 per cent. Thus, the tank which has been fired upon is the only one whose PSD will have a dc component in its output.

The output of the PSD is put through a low-pass filter to remove any ac components and thereby facilitate recognition of the dc component. If any dc component is present, it will actuate a relay which in turn energizes the igniter of a smoke bomb. An additional contact on the relay is available for any external use such as disabling the tank.

Summary

Although the system which has been described

was designed specifically for the tank versus tank gunnery situation, it is pointed out that it has application to many other gunnery situations. While the use of Scotchlite alone as a target retro-reflector limits the maximum range to something under a mile, the use of high quality cataphotes in combination with higher energy arc discharge sources permit operation of a system based on the optical radar principles to be extended to the order of 10 miles. It is also felt that it is important to reiterate the fact that the infrared hit communications system described herein could serve as a complete hit scoring system in an application which did not require evaluation of range estimate.

Contributing Personnel

The work leading to the results presented in this paper was the combined effort of many people. However, the most substantial contributions to the program were made by Dr. H. T. Betz, Senior Physicist, Mr. S. H. Cameron, Senior Scientist, and Mr. E. F. Greta, Research Engineer.

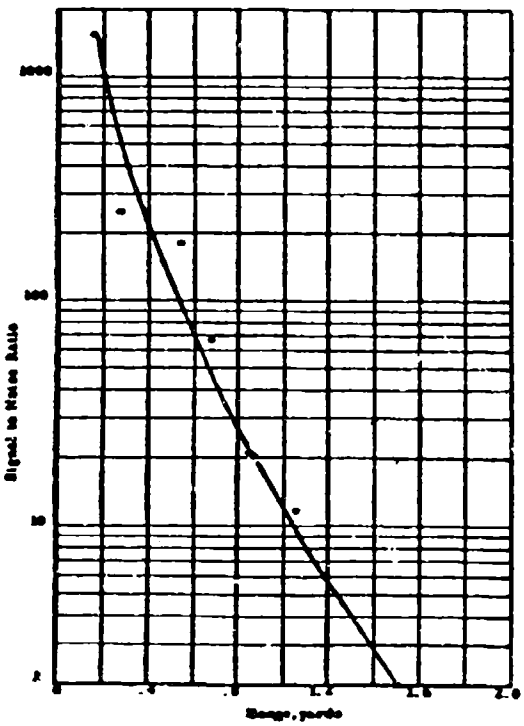


Fig. 2

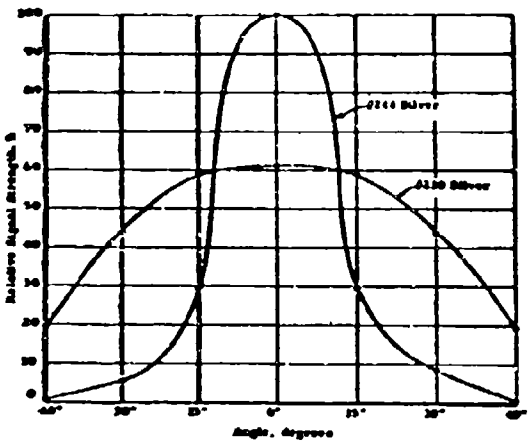
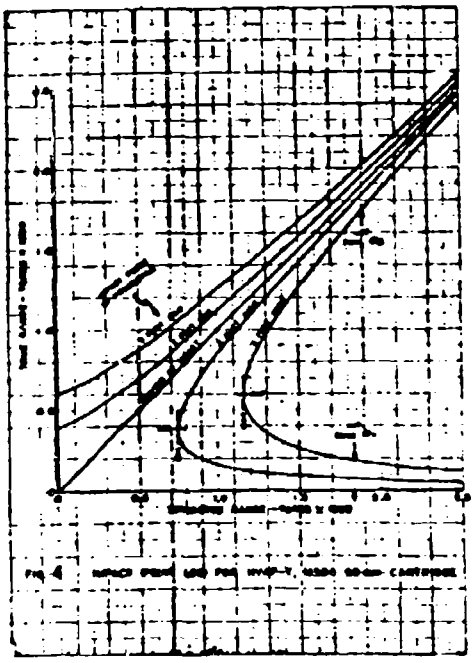
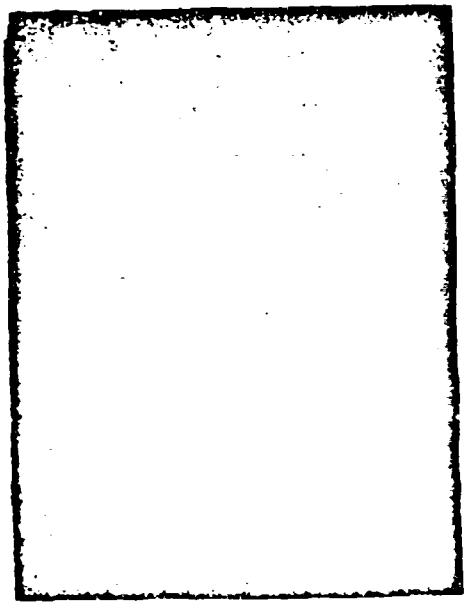
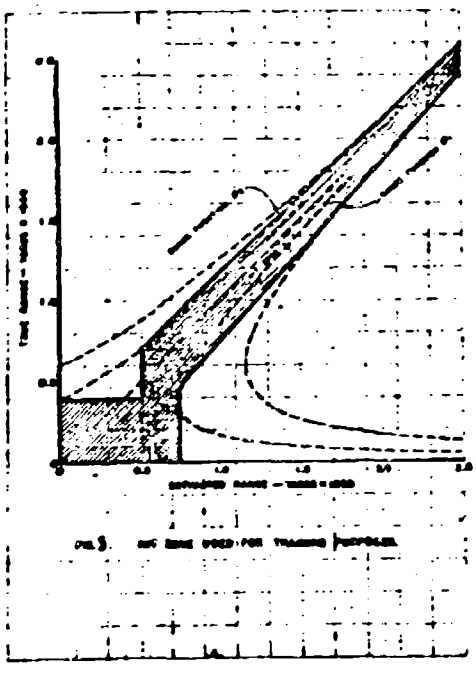


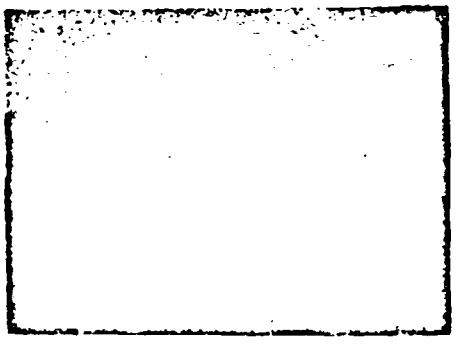
Fig. 1 RELATIVE SIGNAL STRENGTH AS A FUNCTION OF ANGLE BETWEEN NORMAL TO SURFACE AND LINE OF SIGHT

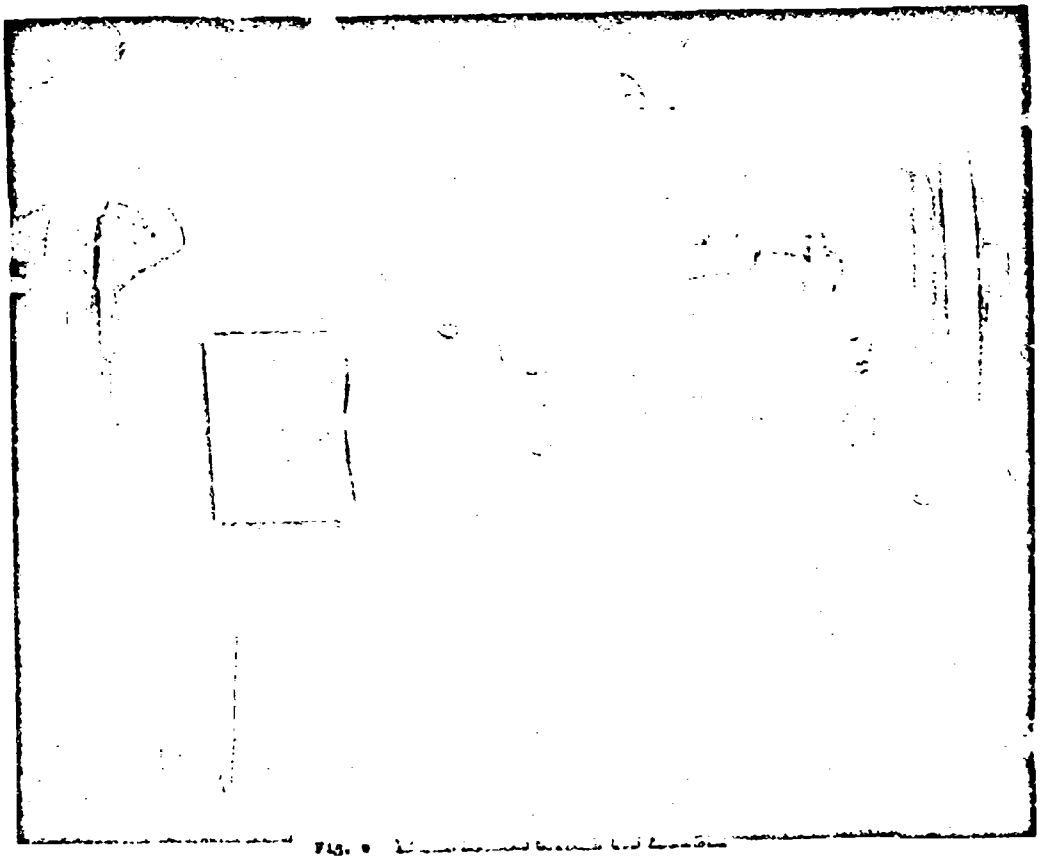
FIG. 3 DETERMINATION OF HIT ZONES FROM PROJECTILE TRAJECTORY.

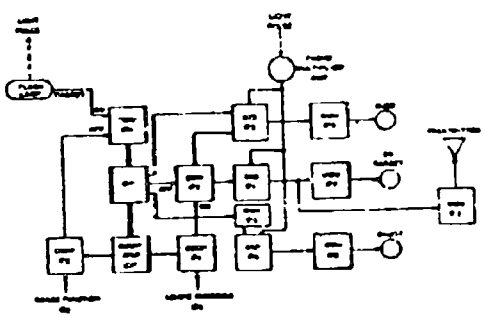












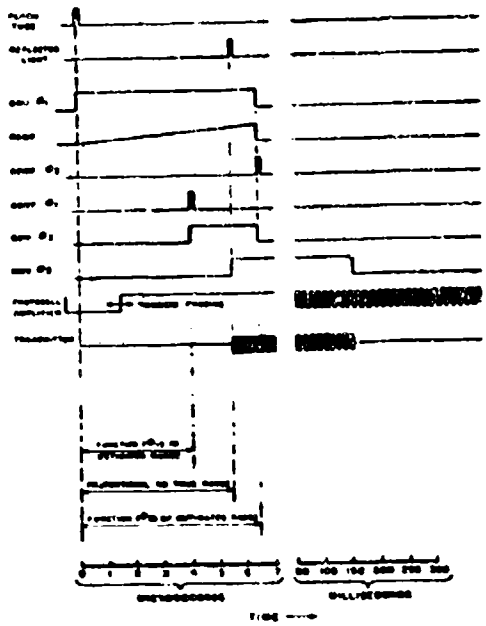
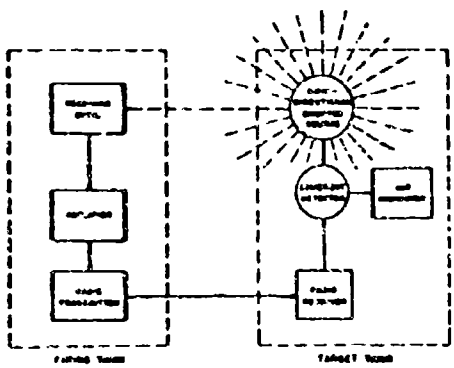
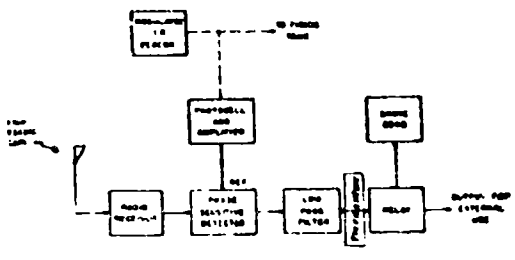


Fig. 10 TIMING RELATIONS IN HIT VERNIER AND TRANSMITTER





MATHEMATICAL MODELS OF MULTIPLE GIMBAL SYSTEMS

BY: Dr. Asriel Rosenfeld
Budd Levyt Electronics, Incorporated

In conventional inertial navigation and guidance systems, some form of stabilized member or "stable platform" is used as a source of navigational data. This platform is maintained in a fixed or known spatial attitude relative to a reference coordinate frame, which may be fixed relative to inertial space or controlled relative to a local vertical. The attitude of the vehicle which is being inertially navigated or guided may then be determined relative to this reference coordinate frame by direct measurement of the attitude of the stable platform relative to the vehicle body axes. In the case of a highly manuevrable vehicle, this attitude may vary arbitrarily. To give the stable platform complete freedom of attitude relative to the vehicle, a multiple gimbal mounting is generally used. This mounting involves at least three gimbals (one per degree of freedom), one within the other.

A recent trend in the inertial systems field is that of keeping track of the vehicle's attitude computationally, rather than by measurement of physical gimbal angles in an actual stabilized platform. In this type of system, which has been variously called "gimballess", "no gimbal", and "strapped down", a computer, usually digital, operates on the outputs of vehicle-fixed sensors in such a way as to continually compute the components of vehicle attitude. One form which these components may take is that of Eulerian angles - in other words, that of the gimbal angles which would have existed had a stabilized platform employing a certain three-gimbal configuration been used.

The "gimballess" approach may be contrasted with the conventional stable platform approach as follows: When a physical platform is used, the gyroscopes or other stabilizing elements which the platform carries make the platform resist any attempt to change its attitude relative to the reference frame. When an angular motion relative to the reference frame is experienced by the vehicle, the platform thus maintains its attitude relative to this frame, and consequently changes its attitude relative to the vehicle. The resultant rotation of the gimbals is thus an indicator of the vehicle's attitude. In the "gimballess" approach, on the other hand, changes in the vehicle's attitude are sensed by vehicle-fixed rotation sensors such as rate gyros. The sensed components of the vehicle's attitude change are used as inputs to the computer which instruments the "phantom stabilized platform". The computer uses these inputs, together with the initial values of the computed gimbal angles, to determine the gimbal angle rates of change resulting from the new inputs, and it integrates these rates to up-date the computed gimbal angles.

The trend to gimballess instrumentation of inertial navigation and guidance systems is thus seen to provide a novel motivation for the mathematical simulation of multiple gimbal systems. The author's interest in this problem, in fact, was aroused in connection with a program of gimballess inertial systems analysis with which he was associated some time ago at Ford Instrument Company. Incidentally, Ford Instrument Company's gimballess inertial system does not simulate a stabilized platform and does not compute Eulerian angles. This system, details of which are classified, thus bears little or no resemblance to the subject matter of the present paper.

In this paper, simplified mathematical models of multiple gimbal systems will be discussed. Equations describing the relationship between the angular inputs and the gimbal angles will be written, and some of the problems involved in the analytical solution of these equations will be discussed.

To begin the discussion, the case of a three gimbal configuration, corresponding to a three axis stabilized platform, will be considered. A typical three gimbal configuration is shown in Figure 1. (The labeling of the gimbals as R, P, and H may be thought of as corresponding to the physical angles "roll", "pitch", and "heading") If Cartesian coordinate systems are imagined to be attached to the vehicle and to the successive gimbal rings, as indicated in Figure 1, then the relationships between these successive coordinate systems for arbitrary values of the three gimbal angles are summarized in Figure 2.

Suppose now that the vehicle is subjected to an angular velocity whose vehicle body axis components as functions of time are known. In order for the platform carried by the three gimbal configuration to remain stable under this angular motion, it is necessary that each of these components of the imposed angular velocity vector be exactly canceled out by the sum of the corresponding components of the gimbal angle rates. Using the coordinate conversion relations of Figure 2, these gimbal angle rates may be resolved into vehicle body axis components. When this is done, the conditions just described may be written as relations between the measured imposed angular rate components, the gimbal angle rates, and the current values of the gimbal angles. These relationships are shown in Figure 3. The equations of Figure 3 thus afford a mathematical simulation of the behavior of the three gimbal configurations of Figure 1 under an arbitrary angular rate input. By solving this system of simultaneous differential equations for the gimbal angles R, P and H, the physical behavior of a three gimbal

stabilized platform can thus be simulated.

As those familiar with the design of stabilized platforms well know, however, there exist situations in which the system of equations shown in Figure 3 cannot be solved. Suppose, for example, that a stabilized platform has gotten into a situation in which the cosine of the gimbal angle P is equal to zero. With reference to the geometry of Figure 1, it is seen that this situation corresponds to a P gimbal angle of $\pm 90°$, which has the effect of making all three gimbal axes coplanar. This situation is commonly known as one of gimbal lock. In the case of a physical platform, when this situation occurs, the platform becomes incapable of compensating for any component of angular rate about the line perpendicular to the common plane of the three gimbal axes. The stability of the platform can thus no longer be maintained in general. A similar situation holds for the phantom platform represented by the equations of Figure 3. When the cosine of the gimbal angle P is equal to zero, these equations suffer what might be called "mathematical gimbal lock". In fact, the equations become as shown in Figure 4. It will be noted that the last two equations of Figure 4 are in general inconsistent, while the first one is ambiguous. The mathematically simulated three gimbal configuration thus suffers from the same failings as does the corresponding physical stabilized platform.

To avoid the gimbal lock problem inherent in three gimbal configurations, stable platform designers have found it expedient to add a fourth gimbal. A typical four gimbal configuration is shown in Figure 5. In the normal position, with all gimbal angles zero as shown in this figure, the innermost gimbal (labeled I) is collinear with the outermost or R gimbal. If coordinate systems are once again attached to the gimbal rings as shown, the needed coordinate relations become those shown in Figure 2 together with an additional one, as seen in Figure 6. The requirements for a four gimbal platform to remain stable under an arbitrary angular input may be described physically in exactly the same manner as was done for the three gimbal configuration; namely, each component of the imposed angular velocity vector must be exactly canceled by the sum of the corresponding components of the four gimbal angle rates. If the coordinate relations of Figure 6 are used to compute these gimbal angle rate components, the stability equations which result are shown in Figure 7.

The addition of the fourth gimbal does not in itself, however, suffice to eliminate the gimbal lock problem. Suppose, in fact, that the cosines of both of the intermediate gimbals P and H are zero. With reference to Figure 5, this corresponds to the physical situation in which all four gimbal axes are coplanar. Clearly, when this happens the physical stabilized platform becomes incapable of responding to imposed angular inputs about an axis perpendicular to the common plane of the gimbal axes. To this physical gimbal lock situation there once again corresponds a mathematical gimbal lock situation. Specifically, the equations of Figure 7 now become as shown in Figure 8, where it is seen that the last two equations are in general inconsistent, while the first one is ambiguous, just as in the case of the three gimbal configuration.

The presence of the fourth gimbal does, however, make it possible to solve the gimbal lock problem by a highly ingenious device. Specifically, it makes it possible to impose a constraint on the gimbal angles while still retaining a sufficient number of degrees of freedom (namely, three) to allow platform stability. Furthermore, it turns out to be possible to formulate this condition in such a way as to prevent the four gimbal axes from ever becoming coplanar, so that gimbal lock becomes impossible. In typical four gimbal configuration designs, the condition in question takes the following form: One of the intermediate gimbals is initially set at the angle zero. Thereafter, the angular rate of this gimbal (or the deviation of its angle from zero, or a combination of both) is measured and is fed back as an error signal by imposing it as an angular rate about some other axis.

For illustrative purposes, it will be supposed that the P gimbal angle is initially set at zero and that some function of its deviation from zero is fed back as an imposed angular rate about the R gimbal axis. This feedback scheme simply adds an error signal term to the first of the equations in Figure 7; the resulting equations are shown in Figure 9. Formally, the equations of Figure 9 are subject to the same mathematical gimbal lock difficulty if the intermediate gimbal angles P and H become equal to $\pm 90°$ simultaneously. However, if the feedback loop is stable and is reasonably tight, the gimbal angle P will never differ greatly from zero, since whenever this angle deviates from zero significantly the imposed error signal will cause one or more of the other gimbals to turn and will thus counteract the effect of the angular rate which is causing the deviation of P from zero. In steady states, the P gimbal axis will thus be perpendicular to the angular rate vector imposed on the platform, so that the angle P will remain equal to zero. If the imposed angular rate vector suddenly shifts, the angle P may temporarily deviate from zero, but the operation of the feedback loop will quickly restore it to the zero position. Evidently, if the angle P always remains at or near zero, and in particular if it never reaches $\pm 90°$, the four gimbal axes can never become coplanar, so that gimbal lock becomes impossible. It is thus seen on physical grounds that the system of simultaneous

differential equations in Figure 9 must always have a solution. The introduction of the fourth gimbal and of the feedback term have thus eliminated the possibility of mathematical gimbal lock.

On the debit side, however, the presence of the fourth gimbal results in a difficulty which did not exist in the case of a three gimbal configuration. This difficulty resides in the fact that the equations of Figure 9, while only three in number, involve four unknowns, the four gimbal angles R, P, H, and I. Clearly, it is not possible to solve this system of three equations for these four unknowns unambiguously. In the physical situation of a four gimbal stabilized platform, the ambiguity in question is of no concern, since physical causes will somehow resolve the ambiguity, and no matter what values of the four gimbal angles result, subject to the equations of Figure 9, they will still correctly determine the attitude of the vehicle relative to the reference coordinate frame as described by the coordinate relations of Figure 6. In the mathematical model of the four gimbal configuration, however, it is necessary to compute explicit, unambiguous values for the four unknowns R, P, H, and I which satisfy the three equations of Figure 9. It is therefore necessary for the mathematical model to include some criteria for resolving the ambiguity involved in solving a system of three equations in four unknowns.

If the equations of Figure 9 were given to a mathematician with no indication of their physical bases, he might immediately suggest that one of the unknowns, for example the innermost gimbal angle rate $\dot{I}$, be made arbitrary, and that the equations then be solved for the remaining unknown gimbal angles in terms of the imposed angular rate inputs and of the arbitrary angular rate of the I gimbal. When the physical background of the equations is considered, however, it is seen that this approach to resolving their ambiguity is physically meaningless, since making one of the gimbal angle rates (whence, one of the angles) arbitrary is tantamount to eliminating the gimbal in question completely.

To take a more physically meaningful approach to the solution of the ambiguity problem, the following alternative reasoning might be considered: Suppose for simplicity that the ambiguity in question is of a one dimensional nature - that is, that it involves two collinear gimbal axes, as was the case in the gimbal lock situations of Figures 4 and 3. A reasonable criterion for resolving the ambiguity can then be formulated in fairly simple terms. For example, if it is assumed that the coefficients of friction and stiction in the two collinear gimbal axes are the same, it is plausible that the inner of the two will respond to the imposed angular rate by turning, but that the outer will remain fixed, since the outer one carries more weight and is thus subject to a greater frictional force than is the inner. In a more general situation, where the ambiguity is not one dimensional, analogous criteria for resolving it could similarly be formulated on a physical basis. These would, of course, be considerably more complicated than those for the one dimensional ambiguity.

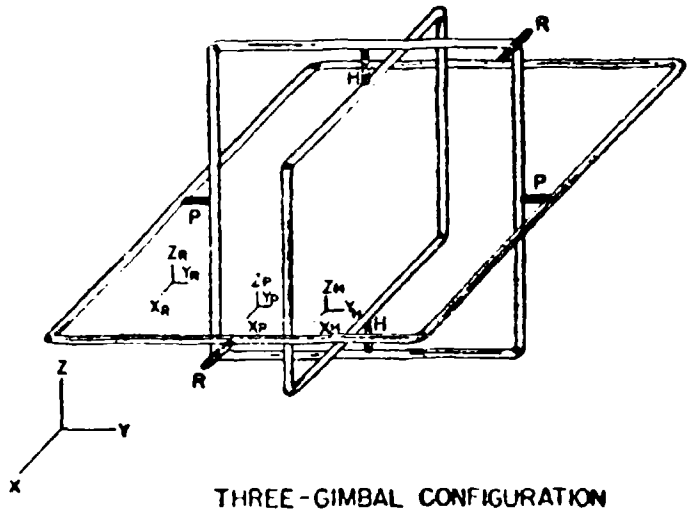
If simulation of a physical four gimbal platform were in question, the formulation of such physically based criteria for resolving the ambiguity would certainly be unavoidable. For the gimballess inertial system application, however, a somewhat simpler approach can be taken, in which the behavior of the platform servo loop is idealized. It may be supposed, specifically, that the servo loop is very stable and very fast compared to the imposed angular inputs. It follows from this assumption that the gimbal angle P will always be zero or very close to it. This in itself, however, does not resolve the ambiguity in the equations of Figure 9, since while the gimbal angle P may itself always remain approximately zero, its time rate of change will not remain zero. (In fact, if both the angle itself and its time rate of change did stay essentially at zero, the P gimbal would essentially be eliminated from the system.) The equations on Figure 9 thus still involve all four unknowns and cannot be solved without resolving the ambiguity.

Suppose now, however, that the three gimbal axes R, H, and I are not coplanar. It may then be concluded that, as a result of the operation of the idealized P gimbal axis servo loop, any imposed angular rate will be "instantly" taken up by the R, H, and I gimbals, and the P gimbal will actually remain stationary at the zero angle value. In other words, the presence of the idealized servo loop may be regarded as making the P gimbal highly "resistant" to turning, or in a sense as giving it a high coefficient of stiction. It follows that when the R, H, and I gimbal axes are non-coplanar, the P gimbal will tend to remain fixed at zero, so that P is in fact eliminated from the equations of Figure 9. The resulting simplified equations are shown in Figure 10. Since these equations involve only three unknowns, they may be solved unambiguously for R, H and I.

Suppose finally that the R, H, and I gimbal axes are coplanar. From the geometry of Figure 5, it is seen that this is equivalent to assuming that the gimbal angle H is 0° or 180°. Since the angle P is zero too, the equations of Figure 9 thus simplify to those shown in Figure 11. These equations are still ambiguous, but their ambiguity is now only one dimensional, since the

conditions that P = 0° and H = 0° or 180° imply that the gimbal axes R and I are collinear. The ambiguity in the equations of Figure 11 can thus be resolved in a manner which is both straightforward and physically correct by simply stipulating that one of the two collinear gimbals remains fixed, while the other does all the rotating required to offset the corresponding component of the imposed angular rate.

A complete mathematical model for an idealized four gimbal configuration may thus finally be described as consisting of the equations shown in Figures 10 and 11. The computer which instruments this phantom configuration solves the equations of Figure 10 for the gimbal angles except when the solution indicates that the angle H is approaching zero or 180 degrees. If H approaches these values more than transiently, the computer switches over to the equations of Figure 11 combined with a suitable one-dimensional ambiguity resolving criterion - for example, the time derivative of I identically equal to zero - until the angle H moves away from its critical value.



THREE-GIMBAL CONFIGURATION

FIGURE 1

$$\begin{pmatrix} X_R \\ Y_R \\ Z_R \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos R & \sin R \\ 0 & -\sin R & \cos R \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

COORDINATE SYSTEM RELATIONS (THREE-GIMBAL CONFIGURATION)

$$\begin{pmatrix} X_P \\ Y_P \\ Z_P \end{pmatrix} = \begin{pmatrix} \cos P & 0 & -\sin P \\ 0 & 1 & 0 \\ \sin P & 0 & \cos P \end{pmatrix} \begin{pmatrix} X_R \\ Y_R \\ Z_R \end{pmatrix}$$

$$= \begin{pmatrix} \cos P & \sin P \sin R & -\sin P \cos R \\ 0 & \cos R & \sin R \\ \sin P & -\cos P \sin R & \cos P \cos R \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

FIGURE 2

$$-\omega_X = \dot{R} + \dot{H}\sin P$$
$$-\omega_Y = \dot{P}\cos R - \dot{H}\cos P \sin R$$
$$-\omega_Z = \dot{P}\sin R + \dot{H}\cos P \cos R$$

STABILITY EQUATIONS
(THREE-GIMBAL CONFIGURATION)

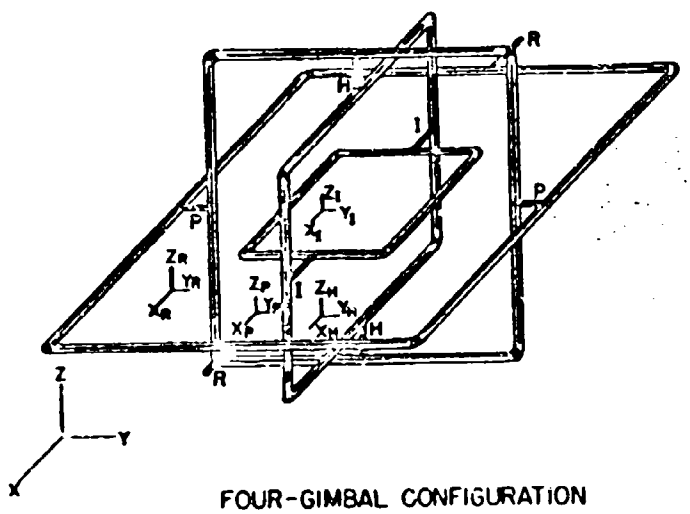
FIGURE 3

$\cos P = 0 \quad \sin P = \pm 1$

$$-\omega_X = \dot{R} \pm \dot{H}$$
$$-\omega_Y = \dot{P}\cos R$$
$$-\omega_Z = \dot{P}\sin R$$

MATHEMATICAL GIMBAL LOCK
(THREE-GIMBAL CONFIGURATION)

FIGURE 4



FOUR-GIMBAL CONFIGURATION

FIGURE 5

SAME AS IN FIGURE 2, TOGETHER WITH

$$\begin{pmatrix} X_H \\ Y_H \\ Z_H \end{pmatrix} = \begin{pmatrix} \cos H & \sin H & 0 \\ -\sin H & \cos H & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_P \\ Y_P \\ Z_P \end{pmatrix}$$

$$= \begin{pmatrix} \cos H \cos P & \cos H \sin P \sin R + \sin H \cos R & -\cos H \sin P \cos R + \sin H \sin R \\ -\sin H \cos P & -\sin H \sin P \sin R + \cos H \cos R & \sin H \sin P \cos R + \cos H \sin R \\ \sin P & -\cos P \sin R & \cos P \cos R \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

COORDINATE SYSTEM RELATIONS
(FOUR-GIMBAL CONFIGURATION)

FIGURE 6

$$-\omega_x = \dot{R} + \dot{H} \sin P + \dot{I} \cos H \cos P$$
$$-\omega_y = \dot{P} \cos R - \dot{H} \cos P \sin R + \dot{I} (\cos H \sin P \sin R + \sin H \cos R)$$
$$-\omega_z = \dot{P} \sin R + \dot{H} \cos P \cos R + \dot{I} (-\cos H \sin P \cos R + \sin H \sin R)$$

STABILITY EQUATIONS
(FOUR-GIMBAL CONFIGURATION)

FIGURE 7

$$\cos P = 0 \qquad \sin P = \pm 1$$
$$\cos H = 0 \qquad \sin H = \oplus 1$$

$$-\omega_x = \dot{R} \pm \dot{H}$$
$$-\omega_y = (\dot{P} \oplus \dot{I}) \cos R$$
$$-\omega_z = (\dot{P} \oplus \dot{I}) \sin R$$

MATHEMATICAL GIMBAL LOCK
(FOUR-GIMBAL CONFIGURATION)

FIGURE 8

$$-\omega_x = \dot{R} + E(P,\dot{P}) + \dot{H}\sin P + \dot{I}\cos H\cos P$$

$$-\omega_y = \dot{P}\cos R - \dot{H}\cos P\sin R + \dot{I}(\cos H\sin P\sin R + \sin H\cos R)$$

$$-\omega_z = \dot{P}\sin R + \dot{H}\cos P\cos R + \dot{I}(-\cos H\sin P\cos R + \sin H\sin R)$$

WHERE $E = 0$ WHEN $P = \dot{P} = 0$

STABILITY EQUATIONS
(FOUR-GIMBAL CONFIGURATION, WITH FEEDBACK)

FIGURE 9

$$P = \dot{P} = 0$$

$$-\omega_x = \dot{R} + \dot{I}\cos H$$

$$-\omega_y = -\dot{H}\sin R + \dot{I}\sin H\cos R$$

$$-\omega_z = \dot{H}\cos R + \dot{I}\sin H\sin R$$

AMBIGUITY RESOLUTION FOR NONCOPLANAR GIMBALS

FIGURE 10

$$P = 0,\ H = 0 \text{ OR } 180^\circ$$

$$\sin P = 0,\ \cos P = 1,\ \sin H = 0,\ \cos H = \pm 1$$

$$-\omega_x = \dot{R} + E(0,\dot{P}) \pm \dot{I}$$

$$-\omega_y = \dot{P}\cos R - \dot{H}\sin R$$

$$-\omega_z = \dot{P}\sin R + \dot{H}\cos R$$

WHERE (e.g.) $\dot{I}$ IS ARBITRARY

AMBIGUITY RESOLUTION FOR COPLANAR GIMBALS

FIGURE 11

OPTIMIZATION OF TEST SYSTEMS

John C. O'Brien, Cooper Development Division, The Marquardt Corporation

The object of this discussion is the development of a method for optimizing testing systems, their performance characteristics, and design philosophy. A testing system consists of a measurement subsystem and a data analysis subsystem. The measurement subsystem is a species of information transmitting system. Its important features are the quantity and quality of information handled. The data analysis subsystem, on the other hand, is a form of decoder. Its important features are the ability to recognize significant information, establish compatibility with a priori knowledge, and refine the signal quality. Its efficacy depends on the form factor or tropy of the signal. A satisfactory choice of system characteristics requires a criterion or set of criteria. The development of these criteria and the appropriate weighting factors, and of their influence on overall testing efficiency is the purpose of this analysis.

Accuracy

The design of an optimum measurement subsystem will be governed by attainable accuracy, in terms of the measurement errors, and by its refinement in the data processing subsystem. It may appear that the total error in an uncorrelated parallel distributed system is lower than in a correlated serial concentrated system. In practice, however, neither a system of completely independent subsystems nor a completely integrated single system are realizable.

Data handling system capacity depends on the number and quality of independent channels. Refinement of overall quality can be achieved in the data processing subsystem by censoring the measurement data, utilizing feedback to minimize internal errors. The ideal data processing subsystem will provide the maximum significance per bit of input information, at minimum cost.

The best testing system will optimize the combination of measurement and data processing subsystems for the maximum output significance, from a minimum of measurements of a given physical situation, in terms of overall cost.

Mathematical Model

To maximize system performance, a mathematical model of the measurement problem is useful. The best obtainable model which describes the measurable energy functions of the physical world is a compromise with the unknown. A composite expression constructed empirically is practical as a first approximation. It should contain the known pertinent characteristics of the physical situation measured, with interrelationships between recognizable states, as well as the performance functions of data processing.

The translation of information from point to point with minimum equipment can be analyzed by Shannon's information channel capacity formula, making the variables functions of both space and time: $C = W \log_2(S+N)/N$. The substitution of space for time is defensible because this law is based on probability theory, which is independent of dimensional specificity. Then "C" becomes a generalized density gradient of information. "W" represents resolution range, "S" is measurement energy extracted, and "N" is non-relevant energy accidentally introduced into the signal channel. Signal-to-noise and accuracy-to-error are two aspects of the same quality factor, and bandwidth is just a special case of resolution spread. Total information translated can be calculated by integrating the product of "C" and the function of space-time which describes the measuring transducer configuration.

The quality of information can be refined by several methods, depending on knowledge of the space-time distribution patterns of signal energy and error/noise energy. Use of this knowledge to modify "W", the resolution function, makes possible including both measurement and data processing in the same equation, and to maximize "C", by the design of static or dynamic space-time filters. The type of a priori knowledge available determines the kind and degree of space-time filtering used.

Representing the pattern of information distribution in space-time by "X" and the corresponding filtering function by $f(x)$ then results in

$$C(X) = f(x)W \log_2 \frac{S+N}{N}$$

Significance

"X" has the character of significance, which makes the data meaningful, in terms of predictable or recognizable patterns. It is the parameter which the data processing subsystem must maximize (at minimum cost) for best performance. $f(x)$ is the transfer function which refines the raw data in accordance with predicted characteristics, by maximizing $C(X)$ and making the data suitable for human consumption as meaningful intelligence. When the Fourier analyticity of the signal can be predicted, auto- or cross-correlation in terms of periodicity can be employed. There is a conflict between the monotony of Fourier harmonics and the variability of information, so care must be taken to prevent smoothing out the information while smoothing out the noise. The signal-to-noise enhancing power of statistical correlation is limited by the extent of periodicity of the signal. With a complex code, 20 to 30 DB improvements in S + N have been obtained by this method. Similar improvements are possible when coding or coherent character of the information can be predicted, if the information extends through a spatial dimension. The concept of spatial Fourier harmonics of physical irregularities, extending

---

"Negentropy" is a double negative

through space to infinity, in all directions, is just as valid as for time functions, and equally useful.

The fact that 20 to 50 DB improvement in accuracy, representing amplitude factors up to 30 to 1, may be realized by "space/time autocorrelation or crosscorrelation" techniques, is not generally appreciated by system designers. A common situation in test system is the specification of extremely high accuracy for energy-sampling transducers as end instruments, assuming that the output accuracy (of the system) will always be lower than that of the input transducers. The possibility of upgrading the input accuracy by 10 to 30 times, allowing the use of inaccurate transducers, and/or enhancing system output accuracies by the same factor, has not been fully exploited.

It is necessary to define a common element for all varieties of signal-to-noise or accuracy-to-error refinement methods. This can be dependence on structural coherence, by which a group of data represents more significance than results in the simple sum of its parts. The term "meaning" is perhaps as suitable as any other for this purpose, and will be used in the following discussion. It depends for its recognition on the intelligence of the designer, and of that part of it which he builds in physical form into the data processing system. This will include both the specialized characteristics for positive signal recognition and acquisition, and the more generalized protective means to reject undesired energy. In the subsequent analysis, the former will be denoted by I_{qc}, and the latter by I_{qn}.

A Priori Information

The necessary a priori knowledge for supporting the filtering function will be denoted by K_{ap}, and includes long-, medium-, and short-term knowledge. In the formula below, this factor with I_{qn} and I_{qc} replaces $f(x)$, so that "W" becomes merely scalar resolution range, which, when optimized, will adequately resolve the existing maximum information density gradient.

Final Model

The mathematical model then becomes:

$$M = WK_{ap}I_{qc}I_{qn}\log_2\frac{S+N}{N}$$

where M "meaning" is equivalent to C(X). This function can be maximized in terms of significant information "M", if "S" and "N" are expanded as dimensional functions of the input situation, and the product $WK_{ap}I_{qc}I_{qn}$, which represents the data processing function, is maximized for a given set of constraints, such as maximum cost.

Complexity

(a) Correlation

The individual transfer function of each of "n" subassemblies in the measurement and data processing systems acts as a filter or correlator on its portion of the input signal to obtain the output signal, and as a crosscorrelator with other subassemblies. These interrelations can be generalized in the form "exp(cqn)" where "c" is an undefined function of correlation and "q" is a generalized dimension.

(b) Reliability

Another result of complexity is the effect on reliability, which is the main ingredient of "I_{qn}", and is also an exponential function of the number of components, as well as of the risk (ratio between component failure and system failure). It can be generalized to $\exp(-xn_i)$, where "x" is the risk factor. Consequently, the effect of the number of parts in a system on the efficiency is proportional to $\exp[(c-x)n]$. Unless "c" exceeds "x", an increase in complexity will reduce the system efficiency.

Gain Distribution

An important aspect of system complexity is the distribution, in space-time, of signal gain and attenuation, versus the distribution of noise/error energy, and the effect on output S/N. The conclusions drawn from a gain distribution analysis made elsewhere are: (1) Gain distribution should be optimized in terms of prevailing noise or error level following each section, rather than on signal level preceding each section (as in conventional AGC); (2) In a system with uniformly distributed noise, the optimum gain distribution will provide exponentially changing gain, progressing from system input to output, for minimum difference between output levels of noise from each source or section of the sequence.

Filtering

The transfer function "$I_{qn}I_{qc}K_{ap}$", modifying the resolving power "W" as a scalar function of time and space (bandwidth), is designed to filter the expected or predicted information, by extrapolating a priori information from known regions of performance to predict performance in unknown adjacent regions in time or space. Among the a priori information available to the designer are the general equations of state, and incidental knowledge of operating conditions and test plans, as well as the statistical nature of the typical noise and error characteristics.

A Priori Information

The utilization function of a priori information, "K_{ap}", depends basically on the initial amount of available information, and the rate at which it is augmented or modernized by error-correcting data during the measurement period. It depends also on its recency, which is subject to degradation with time. The cost of a priori information is low if it can be obtained slowly, but its degradation rate then becomes important. It can be considered as the output of additional communication channels, with narrow bandwidths, using complex coding (mathematical and engineering

language) and very long operating times.

Optimum Performance

The performance of a system as described will be evaluated first by considering that all factors contribute in direct proportion to their individual costs. Then

$$H = I_{qn} I_{qc} I_{sp} W \log_2\left[\frac{S_i}{N_i} f(g) + 1\right]$$

where $f(g)$ is the effective gain function converting S_i/N_i to S_o/N_o.

complexity, expressed in the number of parts, produces the followings

Since $I_{qn} I_{qc} = k \exp\left[\frac{n}{c}(c - s)\right]$

("c" here is correlation efficiency, not cost.)

$$\frac{dH}{dc} = k \frac{n}{c} H$$

Thus the increase in information rate or density provided by an improvement in system integration or correlation efficiency is directly dependent on the number of parts, since except at $S_o/N_o = 0$

$$\frac{dH}{dc} = \frac{\partial H}{\partial I_{qn}} \frac{dI_{qn}}{dc} + \frac{\partial H}{\partial I_{qc}} \frac{dI_{qc}}{dc} + \frac{\partial H}{\partial I_{sp}} \frac{dI_{sp}}{dc} + \frac{\partial H}{\partial S_i} \frac{dS_i}{dc} + \frac{\partial H}{\partial N_i} \frac{dN_i}{dc} + \frac{\partial H}{\partial W} \frac{dW}{dc} + \frac{\partial H}{\partial f(g)} \frac{df(g)}{dc}$$

and, since $\frac{dI}{dc}$ is unity by assumption, for all factors,

$$dH = \frac{\partial H dI_{qn}}{\partial I_{qn}} + \frac{\partial H dI_{qc}}{\partial I_{qc}} + \frac{\partial H dI_{sp}}{\partial I_{sp}} + \frac{\partial H dS_i}{\partial S_i} + \frac{\partial H dN_i}{\partial N_i} + \frac{\partial H}{\partial W} dW + \frac{\partial H}{\partial f(g)} df(g)$$

$$\frac{\partial H}{\partial I_{qn}} = \frac{H}{I_{qn}}, \quad \frac{\partial H}{\partial I_{qc}} = \frac{H}{I_{qc}}, \quad \frac{\partial H}{\partial I_{sp}} = \frac{H}{I_{sp}}, \quad \frac{\partial H}{\partial W} = \frac{H}{W},$$

$$\frac{\partial H}{\partial S_i} = \frac{1.4 H \frac{f(g)}{N_i}}{\left[\frac{S_i}{N_i} f(g) + 1\right]\left[\log \frac{S_i}{N_i} f(g) + 1\right]}$$

$$\frac{\partial H}{\partial N_i} = \frac{1.4 H f(g) \frac{S_i}{N_i^2}}{\left[\frac{S_i}{N_i} f(g) + 1\right] \log\left[\frac{S_i}{N_i} f(g) + 1\right]}$$

$$\frac{\partial H}{\partial S_i} = \frac{1.4 H \frac{S_i}{N_i}}{\left[\frac{S_i}{N_i} f(g) + 1\right] \log\left[\frac{S_i}{N_i} f(g) + 1\right]}$$

Thus, the information density of the system, converted to meaningful intelligence, depends linearly on I_{qn}, I_{qc}, I_{sp} and W, but only reciprocally in rate (i.e., "leveling off logarithmically") with S_i and $f(g)$, and, negatively, and reciprocally with N_i. Since $S_o/N_o = [(S_i/N_o) f(g)]$ is commonly used to describe performance, then

the log function of signal-to-noise or accuracy will increase less rapidly than the other factors, it is evident that the most effective factors in performance are NOT ACCURACY, NOR SIGNAL-TO-NOISE RATIO, BUT INSTEAD RELIABILITY, RESOLUTION RANGE, AND SIGNAL CORRELATION WITH AVAILABLE A PRIORI INFORMATION. Although accuracy and signal-to-noise are necessary elements, their importance decreases as the signal-to-noise or accuracy increases. The improvements in information quantity obtained by achieving greater accuracy or quieter signals can generally be obtained more cheaply by increasing reliability, resolution range, and correlation efficiency (in the use of a priori information) for any signal or measurement level appreciably greater than zero.

$$\frac{\partial H}{\partial \frac{S_o}{N_o}} = \frac{dH}{d\frac{S_i}{N_i} f(g)} = \frac{1.4 H}{\left[\frac{S_i}{N_i} f(g) + 1\right] \log_2\left[\frac{S_i}{N_i} f(g) + 1\right]}$$

The value of H due to S, and to the ratio S/N, increases with increasing information content received. Also, the gain distribution $f(g)$ has a positive effect on the value of H. Differentiation of H with respect to "intelligence" factor "C", in terms of similar dependence on system

Cost

The relation between S (signal power) and its cost is $ for ground-based equipment follows a square law approximately, and includes a factor proportional to frequency.

$$\$_{ec} = K_1 S^2 \sum_{r=1}^{k} \sum_{i=1}^{r} \sigma_i^2$$

Then the equipment is airborne, because of weight problems, and the need for airborne power supplies or for power levels above 10^4 watts, this becomes more nearly a cube law.

$$\$_{ea} = K_1 S^3 \prod_{r=1}^{k} \sigma_i^3 \sum_{s=1}^{n-k} \prod_{i=r}^{s} \sigma_i^3$$

Transporting the equipment into free space beyond the atmosphere or other unusual environments, such as subterranean or submarine regions, is very costly, and for this situation, the cost is estimated to be proportional to the fourth or fifth power, at least, of the actual power level.

$$\$_{es} = K_1 S^5 \prod_{i=1}^{n} \sigma_i^5 \sum_{i=1}^{n-n} \prod_{i=s}^{t} \sigma_i^5$$

Since a high value of $(S+N)/N$ will contribute to the desired maximum information transfer, it is desirable to minimize N wherever possible. The cost of reducing noise and/or error in the measurement system depends inversely on the noise or error power level, and on the type of its source. It appears, from experience, that noise reduction in general costs according to an exponential law,

$$\$_N = K \frac{I^2}{N}$$

Applying the results of these equations to the evaluation of $(S+N)/N$, the following expression is obtained, by including the gain distribution functions:

$$\frac{S_o + N_o}{N_o} = \frac{S_i \prod_{i=1}^{n} \sigma_i}{\sum_{r=1}^{n} N_r \prod_{i=r}^{n} \sigma_i} + 1$$

It appears to be a sound expedient to use $\$_N = AN$, where A is a constant, for conventional resolution ranges in frequency and distance, and A^q for airborne or space-borne equipment (q is a generalized dimension), or for extremely long time spans. Hence

$$\$_N = A^q W I_{qc} \quad \text{or} \quad W = \$_N / A^q I_{qc}$$

The cost function for K_{ap}, the utilization of a priori information, is more difficult to define, because of the various types of information and memory or storage processes which enter into this factor. The first type can be considered as that affecting the quality of design; and the second, the specified performance characteristics. A third type of a priori information includes corrective signals derived from preceding test data which are used to refine or adjust the performance to a limited degree. All of the processes used require prediction, since the benefits of a priori information depend on the technique of minimizing a future error. A reasonable estimate of the cost functions of these three types of a priori information application will then be, for ground based equipment:

Type 1, $\$_1 = K_1$ where $K_{ap} = K_1 + K_2 + K_3$

Type 2, $\$_2 = (K_2)^{p_2}$ where $1 < p_2 < 2$

Type 3, $\$_3 = (K_3)^{p_3}$ where $2 < p_3 < 4$

I_{qm} represents that part of design which is not closely concerned with the particular characteristics of measurement engineering, but rather with more general operating requirements, such as weight reduction, power consumption, environment proofing, human engineering, logistics, maintainability, reliability, stability, miniaturizing, etc., which characterize the areas of application of "good engineering practice." The benefits from the use of the best level of such engineering available are well known; the cost of the benefits is estimated to have higher than a linear proportion to the quality, such as an exponential law.

$$\$_{I_{qm}} = K_I (I_{qm})^{bq}$$

consequently,

$$I_{qm} = \sqrt[bq]{\frac{\$_{I_{im}}}{K_{I_m}}}$$

Since I_{qc} is related to the complexity of the system, when properly applied as $\exp(cqn)$, where c is the system integration correlation factor, then its cost will be derived from

$$I_{qc} = K_{Ic} \left[\exp(\$_{cnq})\right]$$

Hence

$$\frac{dI_{qc}}{d\$_{Ic}} = nq K_{Ic} \exp(\$_{Ic} nq)$$

Recapitulating:

$$N = I_{qm} I_{qc} K_{ap} W \log\left[\frac{S}{N} f(\varepsilon) + 1\right]$$

where

$$f(\varepsilon) = \varepsilon_g \varepsilon_n \varepsilon_o$$

It is now possible to evaluate the entire system in terms of cost "\$", inserting gain factors, and normalizing "S" and "N".

$$\frac{dM}{d\$} = \frac{\partial M}{\partial I_{qm}}\frac{dI_{qm}}{d\$_{im}} + \frac{\partial M}{\partial I_{qa}}\frac{dI_{qa}}{d\$_{ia}} + \frac{\partial M}{\partial K_{ap}}\frac{dK_{ap}}{d\$_{ap}} + \frac{\partial M}{\partial \epsilon_{SS}}\frac{d\epsilon_{SS}}{d\$_{SS}} + \frac{\partial M}{\partial W}\frac{dW}{d\$_W} + \frac{\partial M}{\partial \epsilon_m}\frac{d\epsilon_m}{d\$_{gm}} + \frac{\partial M}{\partial \epsilon_a}\frac{d\epsilon_a}{d\$_{ga}} + \frac{\partial M}{\partial \epsilon_s}\frac{d\epsilon_s}{d\$_{gs}}$$

where "g", "a" and "s" as subscripts denote ground-based, airborne, and spaceborne, respectively. Then, from preceding analyses, the individual improvement factors or profitability are: (normalizing M = I)

$$\frac{\partial M}{\partial I_{qm}}\frac{dI_{qm}}{d\$_{im}} = \frac{(\$_m)^{(1/m_q - 1)}}{I_{qm}\left[K_{im}^{(1/m_q -)}\right]^{m_q}}$$

This indicates that performance value of standard engineering is highest when the initial amounts of engineering is zero, and the initial complexity of the system is small. It asymptotically levels off decreasingly as $m_q \geq 1$.

$$\frac{\partial M}{\partial I_{qc}}\frac{dI_{qc}}{d\$_{ic}} = m_q k_{ic}(\$_{ic} m_q) = m_q K_{ic} I_{qc}$$

This shows that the performance value of system integration/signal correlation is directly proportional to the complexity and the dimensional factor "mq" and increases exponentially with the amount invested.

$$\frac{\partial M}{\partial K_{ap}}\frac{dK_{ap}}{d\$_{ap}} = \frac{1}{K_{ap}}\left[\frac{1}{K_1} + \frac{1}{P_2}\left(\frac{\$_2}{k_2}\right)^{(1/P_2 - 1)} + \frac{1}{P_3}\left(\frac{\$_3}{K_3}\right)^{(1/P_3 - 1)}\right]$$

From this equation it appears that the performance value of a priori information decreases with K_{ap} and also, inversely as the cost exponents of obtaining it.

$$\frac{\partial M}{\partial W}\frac{dW}{d\$_W} = \frac{1}{W K_W^q}$$

This formula denotes the decreasing performance value of resolution range as the amount is increased.

$$\frac{\partial M}{\partial f(R)}\frac{df(R)}{d\$_g} = \left(\frac{\epsilon_a \epsilon_m}{2K_R} + \frac{\epsilon_a \epsilon_o}{5S_{rm}^2} + \frac{\epsilon_a \epsilon_o}{5S^3 r_o^4}\right) \cdot \frac{1}{K_f SN\left[\frac{S}{N} f(R) + 1\right]_{j\,i} \ln\left(\frac{S}{N} f(R) + 1\right)_j}$$

This expression shows how the performance value of (S/N) decreases with increasing S, N, and the entropy or negative information content of the signal.

Since these gradients are all positive, there is no true maximum for (dM/dc). However, it is possible to obtain the second derivatives and determine the maximum incremental profit rate from investment in each factor contributing to the system performance. Only those expressions containing a term in $ will have a second derivative. These lead to:

$$\frac{d}{dc}\left(\frac{\partial M}{\partial I_{qc}}\frac{dI_{qc}}{d\$_{ic}}\right) = (m_q I_{qc})^2 K_{ic}$$

$$\frac{d}{dc}\left(\frac{\partial M}{\partial I_{qa}}\frac{dI_{qa}}{d\$_{ia}}\right) = \frac{\left(\frac{1}{m_q - 1}\right)(\$I_{qm})^{(1/m_q - 2)}}{m_q K I_q I_{qm}}$$

$$\frac{d}{dc}\left(\frac{\partial M}{\partial K_{ap}}\frac{dK_{ap}}{dc}\right) = \left[\frac{1}{P_2}\frac{\frac{1}{P_2} - 1}{K_2}\left(\frac{\$_{P_2}}{K_2}\right)^{(1/P_2 - 2)} + \frac{1}{P_3}\frac{\frac{1}{P_3} - 1}{K_3}\left(\frac{\$_{P_3}}{K_3}\right)^{(1/P_3 - 2)}\right]\frac{1}{K_{ap}}$$

By assuming that all factors have been equalized, insofar as they contribute to M, then the factors of variability are still further simplified.

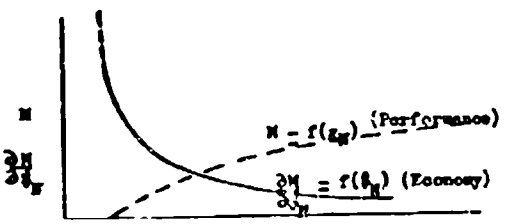
Since correlation plays such an important part in the operation of information handling systems, it is interesting to evaluate the system economy in terms of correlation factor C_c in $I_{qc} = K_{qc} \exp(\$_{Cst})$. Then

$$\frac{\partial M}{\partial I_{qc}}\frac{dI_{qc}}{d\$_{ic}} = \frac{\partial M}{\partial I_{qc}}\frac{\partial I_{qc}}{dc}\frac{dc}{d\$} = \frac{M}{I_{qc}}\frac{n_t \exp(\$_{Cst})}{K_{ic}}$$

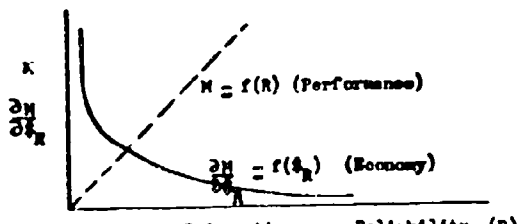
From these expressions and the corresponding plots in figures 1 - 6, the profitability or profit rate from investing money from a limited budget in any of the factors can be determined, and the total investment optimized. Examination of the curves before inserting values from experience in the constants of integration and the fixed coefficients, shows the trends in a broad sense. These include the constant profitability of "W", or resolution range, and of "K_{ap}", a priori information, and of "I_{qc}" for simple systems. The decreasing profitability of power gain, and of the use of noise

reducing design, appears in the corresponding curves. Finally, the exponentially increasing benefits of the use of correlation, where this is justified by the nature of the physical situation and of the significance attached to the desired information, is shown in the "I_{qc}" rate curve.

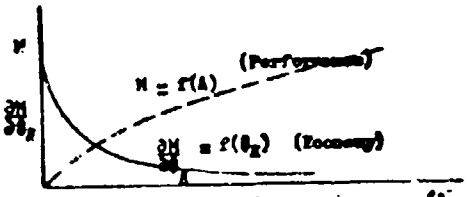
In conclusion, the best investment for improved testing performance is in system integration and in matching the system, as a correlation filter, to the testing problem. The generally prevailing emphasis in contemporary engineering design on greater accuracy and signal-to-noise ratios leads to lower returns in performance per dollar, than if the same emphasis were to be placed on bandwidth, filter design, correlation techniques and overall system integration. This is true regardless of the "state-of-the-art" of these techniques. When the cost of increased accuracy and of better signal-to-noise ratios is considered, the extravagance of specifying the highest accuracy obtainable as a performance "safety factor" is obvious. In short, in most information gathering systems, internal accuracies and signal-to-noise ratios are the least effective contributors to the overall performance. Their importance has been greatly overestimated in the past, and continues to be by a great majority of engineers.



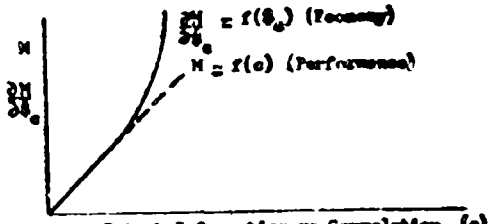
— Output Information vs. Noise Reduction
— Return as function of Cost
Figure 1



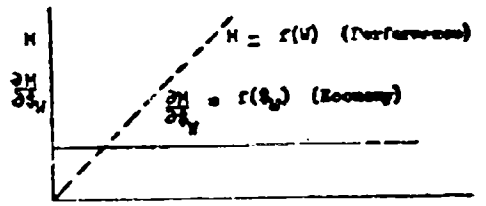
— Output Information vs. Reliability (R)
— Return as function of Cost
Figure 2



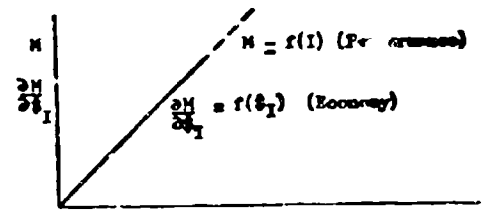
— Output Information vs Accuracy (A)
— Return as a function of Cost
Figure 3



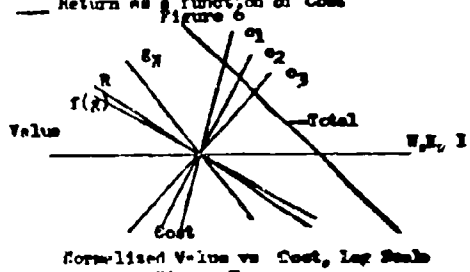
— Output Information vs Correlation (σ)
— Return as a function of Cost
Figure 4



— Output Information vs Resolution (M)
— Return as a function of Cost
Figure 5



— Output Information vs System Integration (I)
— Return as a function of Cost
Figure 6



Normalized Value vs Cost, Log Scale
Figure 7

IMPACT OF THE RECENT INTERNATIONAL RADIO CONFERENCE, GENEVA ON USAF COMMUNICATIONS-ELECTRONICS PROGRAMS

By: Mr. C. V. Lester, Headquarters USAF

The International Radio Conference which met in Geneva, Switzerland, from August until late December of last year gave international recognition for the first time in the history of international radio regulations to such new techniques and services as the space radio service, radio astronomy, and scatter communication. The Conference also took forward-looking steps to enhance the safety of life and property in the air and over the sea by adopting regulations which, among other things, more clearly delineate the frequencies for use in cases of distress and emergency. Other regulations recognize the need for distinguishing between radars which are used for radio navigation, defense, search and storm detection. The Conference recognized the critical situation in the high frequency fixed bands where the accommodation of additional requirements for frequencies is almost impossible. This paper undertakes to examine each of these items with reference to their impact upon Air Force communications-electronics programming.

RADIO NAVIGATION AND RADIO LOCATION

The 1947 Radio Conference lumped all radar uses under "radio navigation" in the Frequency Allocation Table which it adopted. That Conference did not envisage the many uses to which this technique would be applied within the following decade. However, it soon became obvious that it was highly desirable to separate the frequency assignments of radars used for aerial navigation and safety from those of other important types of radar applications, such as defense radars, aircraft anti-collision radars, and storm detection radars.

In consideration of this situation, the United States with strong Air Force support proposed to the 1959 Conference that separate frequency allocations be made for navigation radars as distinguished from other radars. To an unexpected degree, the Conference adopted this concept. It provided that radars be placed into two categories: radio navigation and radio location. It also made separate frequency allocations for each. Thus, in the future, navigation radars will for the first time be protected from interference caused by other radars, while at the same time the important non-navigation radar uses will not be curtailed and hindered by the necessity to protect navigation.

Substantial allocations for both radio navigation and radio location services were made throughout the useable spectrum above 100 Mc/s. These allocations will serve to promote the orderly growth of these services.

SPACE COMMUNICATIONS

The rapid development of radio communication is nowhere more strikingly brought out than in the field of space communication. The Atlantic City International Radio Conference of 1947, which adopted the Radio Frequency Allocation Table currently in force, made no provision for extra-terrestrial communication for the simple reason that only thirteen years ago the need for such communication was not even envisaged by the delegates to that Conference. As a matter of fact, it was not until preparations were begun for the International Geophysical Year that scientists and communicators gave serious consideration to the problems which might be encountered when attempting to communicate with man-made objects in space. Therefore, adoption by the 1959 Radio Conference of specific radio frequency allocations for space communications can rightly be considered as innovation.

The United States and the United Kingdom delegations took the initiative at the Conference to bring a degree of recognition to this new use of radio communication which now occupies international status along with old-line radio services, such as the fixed and mobile services which have been recognized internationally for more than fifty years.

Although the United States insisted at the opening of the Conference that some specific radio frequency bands were necessary immediately for the new space service, many countries, particularly the smaller ones without the capability of extensive space research, were only mildly interested in allocating portions of the already seriously crowded radio frequency spectrum for the space radio service. Even the Soviet Union Delegation maintained throughout the Conference that allocations for space communication would be premature. That Delegation did, however, agree to be a member of the Conference sub-committee which inaugurated the study of the problem and recommended band allocations to the main committee dealing with frequency allocations.

After prolonged discussion, the Conference finally recognized the space radio service by definition and made provision for it to operate in thirteen bands interspersed throughout the radio spectrum above 10 Mc/s. It is interesting to note that the band 136-137 Mc/s is one of the bands which will become available ultimately to the space service on an exclusive basis. This band will accommodate those operations near 108 Mc/s which were associated

with earlier space research. Another important contribution made by the Conference was a recommendation that a special international conference be convened in 1963 to study further the requirements of the space service and to allocate, if necessary, additional frequency bands for this service. It is noteworthy that the Conference recognized the fact that space communication is still in its infancy and that substantial development must take place before the space radio service is firmly established. Accordingly, the Conference provided that the newly allocated bands for the space service shall be used initially for research purposes. This action was taken without prejudice to any ultimate use of these allocations by the space service.

The immediate impact of the Geneva Conference on the military communications and electronics program is to provide on an international basis valuable allocations to satisfy the frequency requirements for space research activities.

The long range and perhaps more important impact may well stem from the Conference recommendation to hold an extraordinary world-wide radio conference by 1963 to consider "the allocation of frequency bands essential for the various categories of space radio communication." This conference can change the frequency allocations to the space service, and these changes could affect the frequency space available for other military operations. This problem is under study in the Interdepartment Radio Advisory Committee.

RADIO ASTRONOMY

As in the case of the space radio service, the radio astronomy service is a late comer in the international radio frequency picture. Accordingly, it is hardly surprising that the Conference was not prepared to give this service the ultimate frequency allocation or spectrum reservation which it deserves.

The United States proposal to the Geneva Radio Conference with respect to radio astronomy was that the band 1400-1427 Mc/s be allocated exclusively on a world-wide basis to radio astronomy. It was not until the radio conference was well under way that scientists and radio astronomers in the United States and Europe realized that this allocation was inadequate for their purposes. In consequence, they sought to have the Conference reserve additional space for astronomy. They were particularly interested in radiations from such substances as deuterium on 327.4 Mc/s, hydrogen on 1420.4 Mc/s, and the hydroxyl radical (OH) on 1667.0 Mc/s. The Conference ultimately made provision on some 23 discrete frequencies or bands for radio astronomy. In only one case, namely the hydrogen line in the band 1400-1427 Mc/s, was a clear allocation made. In other cases, the Conference either recommended that certain bands be kept clear for radio astronomy or, by means of footnotes on the allocation table, provided for the protection of radio astronomy bands. Although allocation status was not generally accorded to radio astronomy, the Conference was very much aware of the necessity for protecting this service. Radio astronomy is already providing much knowledge of the nature of space. Undoubtedly, future radio conferences will undertake to insure greater protection against man-made interference to the radio astronomical observatories which are being established in many places.

SAFETY OF LIFE AND PROPERTY

The United States Air Force Air Rescue Service provides world-wide air rescue service for all Air Force and other U. S. Military activities requesting these services. In addition, the Air Rescue Service renders assistance, upon request, to civil aviation of the United States and to civil and military aviation of other countries. The results of this Conference concerning the regulation of frequencies involved in the safety of life and property have a direct bearing on communications involved in Air Force Air Rescue operations. Accordingly, it was a source of great satisfaction to the Air Force that a number of proposals which the Air Force either made or actively supported were adopted by the Conference. The first of these was to increase the utility of the frequency 8364 kc/s upon which survival craft transmit. While this frequency has since 1947 been recognized as available to survival craft, its use was shared between such craft and ship radio stations. As a consequence, when the low power transmitters of survival craft used this frequency during emergencies, the extremely heavy interference from ship stations precluded searching aircraft and ships from hearing the survival craft. The Radio Conference agreed with the Air Force view that, in the interest of safety, it was necessary to prohibit the use of this frequency by ship stations for routine calling purposes. Accordingly, the frequency is now reserved exclusively for survival craft and other craft engaged in search and rescue activities.

The frequency 243 Mc/s has for some years been used for search and rescue purposes by military aircraft. The Air Force believed that, in the interest of expediting rescue efforts, this frequency should be made available on much the same basis as the frequency 8364 kc/s. The Radio Conference readily agreed with this view and, as a result, there are now two frequencies available to survival craft and are reserved for search and rescue work.

Another step taken by the Conference to enhance the safety of life and property in the air and on the sea was the increased emphasis given to the frequency 2182 Kc/s as the international telephone distress frequency. This frequency was given almost the same degree of protection as that which has historically been afforded 500 kc/s, the international radio telegraph distress frequency. From the standpoint of the Air Force this is important since, in the past, there has not been available a good, reliable frequency within the range of Air Force HF airborne equipment for emergency two-way communication with coast stations and ships. There have been occasions where such a frequency could have been used to an advantage. Frequency 2182 kc/s is now available for this purpose. It will be well guarded by coast stations and ships.

SCATTER COMMUNICATION

United States proposals for radio frequency allocation for ionospheric forward scatter communication between 30 and 50 Mc/s were received by the Conference with mixed reactions. A few countries such as the Soviet Union and Brazil were interested in establishing ionospheric forward scatter circuits and, therefore, tended to support the United States proposals while most of the countries either had no interest in this technique or had fears that increased use of scatter would threaten their national television services. While the United States proposal for ionospheric scatter envisioned employment of frequencies immediately above and below 40 Mc/s so as to be able to avoid F-2 propagation, the countries of the European area, where television service is assigned the band 41-68 Mc/s, objected. After prolonged discussion, the Conference decided to confine ionospheric forward scatter transmission to three 4 kc/s bands roughly between 32 and 39 Mc/s. The Conference adopted a "grandfather clause" which provides that existing scatter operations below 40 Mc/s may continue on their present frequencies provided they do not interfere with other services.

From the standpoint of the Air Force, existing ionospheric scatter circuits are not affected. However, lacking a frequency allocation for ionospheric forward scatter above 40 Mc/s will make it necessary to use present frequency assignments and assignments in the new scatter bands with caution so as to avoid the possibility of interfering with other services because of F-2 mode of propagation.

Tropospheric forward scatter techniques, on the other hand, met with general conference approval. The bands allocated to the fixed service are generally available for tropospheric scatter circuits.

CONGESTION IN THE HIGH FREQUENCY FIXED BANDS

The continually increasing requirements of all countries for high frequencies for long distance and international communication is a matter of concern to the Air Force and to all other users of the high frequency portion of the radio spectrum. The density of frequency assignments to stations and the volume of operations in the available frequency bands are now so great, that the accommodation of new or expanded operations continues to become more and more difficult.

Among the reasons for this situation, the following may be mentioned:

1. The steady increase in the number of independent countries, each of which desires to establish and maintain its own global communications facilities,

2. The expanding economy of many countries with attendant communication requirements,

3. The reduction in international rates and improvement in service, resulting in greater public demand for service,

4. The availability of new customer-to-customer teletype and facsimile services on a global basis, and

5. The inability to provide wide-band submarine cable facilities in sufficient quantity and between enough countries to keep up with the growing demand for service.

The Geneva Radio Conference was well aware of this situation. In an attempt to find a solution, the Conference adopted a series of resolutions and recommendations. The first of these would set up a panel of experts to study the over-all problem of congestion with the view to making specific recommendations for more economical use of this important portion of the radio spectrum and the possible diversion of traffic to other radio frequency bands and other modes of communication. Another conference action requests that administrations reduce the pressure on the high frequency bands by adopting new frequency-saving techniques to the maximum extent possible. Still another recommendation refers to the consolidation of lightly loaded radio circuits. It urges administrations to conclude special arrangements on the common use of existing international radio circuits and recommends that, in planning new radio circuits or the extension of the existing circuits, administrations take into account the Conference recommendations on this subject.

* * * * * * *

In conclusion, it should be pointed out that, from the standpoint of the Air Force, this paper has treated only the more significant changes which the 1959 Radio Conference effected in the International Radio Regulations. These and other changes adopted by the Conference are being reflected in the revisions to the National and Military documents which serve to guide the development of equipment and regulate the operation of communication and electronic systems of the Military Services.

THE USE OF APPARENT FARADAY ROTATIONS IN PREDICTION OF IONOSPHERIC DISTURBANCES

By: Messrs. H.S. Marsh and R. J. Cormier, AF Cambridge Research Center

ABSTRACT

Apparent FARADAY ROTATIONS on WWV and CHU short path signals are studied.

These rotations vanish during large magnetic disturbances, (high Ak index) but have shown characteristic daily pattern changes about 1 day before onset of large disturbances, and other pattern changes during return to quiet conditions, throughout Oct 1959. They are frequently preceded and followed by marked level differences on WWV 440 and 600 cps tones and carrier.

These pattern changes may permit advance prediction of disturbed ionospheric conditions, as well as of impending return to quiet conditions.

Strong correlation with certain forms of Sporadic E has been found.

The rotations appear best during magnetically and ionospherically quiet periods, when normal fading and multiple reflections seem abnormally minimized.

Study is being continued to check morphology and refine the possibility of short-range predictions using this phenomenon, and possible extension to prediction of Arctic Blackouts.

Variations in times of E to F layer or scatter transitions appear linked to the rotation and remodulation changes, but connection is not yet clear, though both mechanisms appear closely allied.

INTRODUCTION

The Propagation Laboratory of ERD, AFCRC, has recorded amplitudes on various WWV and CHU frequencies, at Scituate and Concord, Mass., for about one year, in a study of Ionospheric storminess indices.[8]

During the study it was found that two unusual anomalies occurred irregularly but persistently on certain frequencies, both from Washington, WWV, and from Ottawa, CHU, airline distances to Concord of about 638 and 502km, respectively.

The first noticed is an apparent Faraday rotation, not occurring only when the signal was near the MUF, but often persisting throughout the daylight hours, and often of machine-like regularity, most frequent on CHU 7.35 and WWV 10mcps.

The other anomaly, apparently more frequent in the winter months, is an apparent frequency selectivity of propagation on the two tones and the carrier-only signals of WWV, often causing a clean-out level difference of several db. between the three signals. This phenomenon, originally called 'two-tone remodulation', typically appears in the twilight hours before and after the apparent Faraday rotation, usually while the signal appears to be propagated by a scatter mode, before and after the period of reflective layer support.

The fact that appearance of the apparent Faraday rotations appears to be strongly allied to variations in the magnetic index, and appears to show changes in pattern somewhat in advance of changes in the ground-level A_k values, encouraged further study of both phenomena, leading mainly to upsetting of several explanations which had been advanced for their occurrence.

Like Sporadic E, to certain forms of which it shows a close alliance, the exact origin and philosophy of this closely coupled pair of anomalies is now largely a matter of conjecture.

FORM OF THE ANOMALIES

Fig. 1 shows a sample of the apparent Faraday rotation, on WWV 10mc, recorded simultaneously at Concord and Scituate. Instrumentations are similar, except that Concord uses a 17-foot vertical whip antenna, while Scituate has about 20 feet of flat-top, strung about 15 feet above ground. The calibrations are not identical.

It is seen that the patterns at these two stations, on this magnetically quiet day, are quite similar, but not identical. The A_k value reported by Ft. Belvoir for the day was 4.

As noted on the Scituate record, the low frequency rotations most evident are probably due to interference of the Lower ordinary and extraordinary rays, while occasionally superimposed on them can be seen higher frequency interferences attributed to interference between lower ordinary and Pedersen ordinary rays.[6]

At magnetically disturbed times, this regular, long-lasting phenomenon is usually absent or masked by the random fading which prevents its recognition except for brief periods when near the MUF. The relative infrequency with which it appears on other circuits arouses suspicion that it may be a localized critical effect of the high reflection angle, (short path length), or of a magnetic field alignment sensitivity.

The two receiving stations are separated about 35 miles airline, at nearly the

same distance from Washington, and separated perhaps three degrees in azimuth. As will be shown later, on magnetically disturbed day the two stations' patterns may differ vastly, lending some weight to the suspicion of field alignment sensitivity.

For a rough check, a few days' observations were made at Chilton's Pt., Va., near the edge of the ground-wave range of WWV, and no trace of either anomaly was found. In addition, through the courtesy of Drs. H.G. Booker and S.M. Colbert, of Cornell, a brief examination of WWV 5mc recordings at Ithaca was made. It was not extended enough to be conclusive, but gave no indication of the long-persisting Faraday or two-tone remodulation, for the brief stretch of data inspected.

If the assumed interference factors noted on Fig. 1 are correct, the cause of the rotations is simply explained. What is not clear is their long persistence, sometimes on several frequencies simultaneously, and far removed from the MUF, and the apparently extreme quiescence of the layer conditions separating and reflecting the components with an almost entire absence of the random fading which masks such anomalies on most HF circuits.

Gardner and Pawsey[2] studied such rotations at vertical incidence with pulses at a frequency slightly above the gyro-frequency. They found separation of the ordinary and extraordinary rays apparently occurring at E sub-layers around 73 to 85km high. This rotation would appear morning and evening, and usually disappear around midday, when high ionization would cause complete absorption of the extraordinary ray. The WWV and CHU signal rotations showed the same characteristic midday disappearance during the summer months, which was not present during the winter.

Other evidence[1] indicates the probable existence and importance of such sub-layers. Existence of the Gardner-Pawsey sub-layer structure as a probable generator of the Faraday interference modulation regardless of whether signal reflection occurred at the E or F layer, was tentatively accepted, as a basis for speculation.

This supposition was reinforced by an attempt to correlate occurrence of the rotation with E-layer maximum-total-density. Since Faraday rotation on satellite signals is used as a guide to total electron density in the path, it was assumed that a direct relation between E layer maximum density, as derived from mid-or end-point virtual height and critical frequency soundings, should appear. Instead, for the period sampled, no connection between occurrence of the rotations and the indicated density could be found. This reinforced the concept of a semi-independent sub-layer, following controls not revealed by the gross E layer behaviour; possibly due to ionization stratification, or to amplified magnetic sensitivity.

Fig. 2 shows a sample of the frequency-selective propagation, or "two-tone remodulation", on what is believed to be a ground-scattered signal. The two-minute carrier-only periods, (the lowest line of traces), and the separate level averages of the two three-minute tone groups, can be clearly distinguished. The steady rise in general level until 0415, and its equally steady decline thereafter, might be due to a ground-scatter focussing effect. This type of apparent scatter, with or without visible tone distinction, is quite common at night. Its diurnal pattern, and in particular the absence of the typical '20 hours dip', make it probable that it is ground, rather than classical mid-point ionospheric scatter.

When first observed, the small level differences and their relative attenuations permitted possible interpretation as due to the Luxembourg-like tone self-demodulation or cross-demodulation shown possible for low audio frequencies on carriers near the gyro frequency by Hibberd[3] and others. So it was first called 'two tone demodulation'. Later it was found that the two tones could change their relative amplitudes, and sometimes the (originally weaker) carrier exceed either of them, and, more rarely, level differences of 20db or more developed, so the local name was hurriedly changed to 'remodulation', since all these latter changes are contrary to the mechanics outlined by Hibberd.

WWV on 10mcps transmits the tones as upper sidebands, added to the carrier for three minutes out of every five, alternating the 600 and 440 cps tones, and starting with 600 cps after the hourly five minute noise break, and announcement. Noticeable is the absence of interfering signal during the noise break at 0445 EST.

Several powerful signals from WWV, all modulated at the same tone frequency, are present simultaneously in about the same ionospheric reflection area. Hibberd shows the possibility of a low audio frequency heating effect transferring partially non-recoverable energy to the electron plasma, resulting in attenuation of the upper portion of each cycle of a powerful wave. This may cause attenuation of the order of several db, in addition to the generation of harmonics. The effect may occur either from self-demodulation or cross-demodulation by

another signal. Spectrum analysis has shown the presence of significant second harmonics, and weak but recognizable third harmonics. The relatively vast differences in amplitude between the tones and carrier sometimes seen, as well as the inversion of the expectable relative amplitudes of the tones, are both contrary to the Hibberd self or cross modulation mechanics, but its possible presence has not been ruled out as yet. No attractive other explanation for the phenomenon has been found as yet. One possibility, now being explored, lies in the fact that Concord is slightly under one wavelength distant, by airline, from Washington for one tone, and slightly over one wavelength for the other. If two slowly varying paths of unequal length are possible, a mechanism of differential audio interference patterns could be generated. Such a mechanism, however, would not be compatible with Fig. 2, in which all three components reach an amplitude peak and start weakening at about the same time.

Fig. 3 shows a fairly common transition pattern from daytime F--reflection to nighttime scatter signals. At first, a fairly clear rotation pattern is present. Then after the 1845 EST noise break, a thinned trace characteristic of approach to the MUF[9] appears. A few minutes before reflection fails, around 1920, a few depressions of the carrier-only signal appear. Then the signal drops to the quite steady scatter level. The beautifully separated levels of Fig. 2 are not visible, possibly due to the interfering signal visible at the noise breaks, but quite regular depression of the carrier level during its two minute periods can be seen and measured.

ANOMALY PATTERNS AND THEIR CORRELATIONS

Sample correlations of the anomalies were made with various ionospheric parameters as a preliminary to more detailed study.

Fig. 4 shows the incidence of clearcut rotation on CHU 7.3 mcs through most of Oct 59 compared with Belvoir daily A_k variations. The second line from the top shows the quite regular occurrence of forenoon and afternoon rotations typical of magnetically quiet days for this frequency and season. The left portion of the upper line, on the other hand, shows the complete absence of Faraday typical of disturbed periods. An interesting point here is that on the 4th, while A_k is high, rotation appears in the afternoon, and is followed about a 50% drop in A_k for the following day. While A_k remains moderately high, afternoon Faraday persists through the 7th, ending in the very quiet spell discussed.

Conversely, following a moderately quiet spell, in the 17th afternoon Faraday only appears for a very short period, and the following day the A_k rises to over 30.

These two isolated occurrences do not prove anything, but considered with the long, consistently present or consistently absent times, they hint the possibility of a magnetically sensitive sub-layer, possibly as predicated by Gardner and Pawsey, which might telegraph ahead magnetic field changes substantially before they reached the earth's surface.

If further study confirms and clarifies this apparent effect, it could prove a strong tool toward prediction of magnetic and ionospheric disturbances.

CORRELATION WITH E_s

Three major types of Sporadic E have been observed during the study period at Ft. Belvoir, Va., Ft. Monmouth, N. J. and Ottawa, Canada, but not necessarily simultaneously.

1. Blanketing E_s which does not permit E or F region returns, frequently occurs at around 100 kms, below the E layer maximum height.

2. Transparent E_s which permits F region returns.

3. Weak E_s which permits both E and F region returns.

Good correlation has been found between blanketing E_s and the appearance of apparent Faraday rotation. This type of E_s is a large ionized cloud several kms in area rather than a series of small fast moving clouds of ionization.

The close correlation observed briefly between the Faraday rotation and a certain form of Sporadic E, if the hypotheses that the Faraday rotation takes place at a fairly specific E-sub-layer, is interesting in the light of a recently advanced theory, suggesting the formation of Sporadic E as possibly due to turbulence and vertical wind-sheer arising in the same general sub-layer area.[9]

This possibility is supported by the fact than in the temperate zone, E_s has a negative correlation with magnetic activity, as does the Faraday occurrence.

In Fig. 5 Ottawa to Boston MUF's have been plotted based on Ottawa soundings. The MUF for the blanketing E_s is quite high compared to the E and F MUFs and the Faraday Rotations appear with the occurrence of blanketing E_s as can be seen

from 0600-1300 hours and from 1800-1900 hours. Transparent E_s and weak E_s occurring at other times have shown no correlation with the appearance of the anomaly. More information concerning the probable sublayers of the E region is needed before more definite connections between E_s and apparent Faraday mechanisms can be established.

Fig. 6 shows a worksheet exhibiting signal and anomaly morphology through late Dec 59 and early Jan 60. As a worksheet, it is too crowded for comfort, but shows interesting trends.

On the 28th, with high Storm Index and A_k, only a short patch of afternoon Faraday appears, followed by several hours of heavy "two-tone", after reflective layer failure.

On the 29th, and briefly on the 30th, with decreasing Storm Index and A_k, a rare event occurs. This is the simultaneous occurrence of Faraday and twotone, which appear otherwise almost always mutually exclusive. Thence onward a fairly regular daily pattern of Faraday appears, with a few spells of twotone preceding it, without obvious motivation. The midday Faraday disappearance noted earlier in the fall has vanished.

A possibly significant fact, not illustrated here, was that for about a week, evening twotone always accompanied scatter for several hours when the scatter field was weak, and never accompanied the scatter when the scatter field was strong, particularly shortly after the failure of reflective support.

2 HOP TRANSITION

An interesting fact is that 2 hop propagation is predominant in the evenings at this season, and 1 hop seems vastly attenuated. A simultaneous signal dropout, not shown here, was observed at Concord and Scituate, Mass. Times of this dropout corresponded to the 2 hop mode failure; the signal level did not disappear but a weak signal corresponding to the 2 hop mode, which was possible at the time, remained. No interference appeared at the WWV noise breaks.

In Fig. 6 the time of drops in signal level showing reflective failure during evening hours corresponds to the 2 hop times for every case. Propagation goes from 2 hop to scatter ratner than 2 hop, 1 hop then scatter. Two examples are shown in Table No. 1.

TABLE No. 1

28 December 59

hours	1800	1900	2000	2100	2200
possible mode	1 hop 2 hop	1 hop	1 hop	1 hop	scatter
actual mode	1 hop 2 hop	scat- ter	scat- ter	scat- ter	scat- ter

30 December 59

hours	1800	1900	2000	2100	2200
possible mode	1 hop 2 hop	1 hop 2 hop	1 hop	1 hop	scatter
actual mode	1 hop 2 hop	1 hop 2 hop	scat- ter	scat- ter	scat- ter.

The transition period from scatter to layer propagation at sunrise does not display this 2 hop mode depending but rather layer support begins with the 1 hop mode. The time difference between 1 hop mode and 2 hop mode at sunrise is less than one hour while the evening difference is at least 2 hours.

The example of Jan 1st on Fig. 7 shows a magnetically quiet day, but is abnormal it two respects. First, the 440 cps tone median is greater than the 600 cps tone; second, the db spread between tones and carrier is 20 db greater than normal. The 440 cps tone 600 cps tone median amplitudes do not fluctuate with respect to each other. The sudden rise in signal level at 0745 corresponds to the transition from scatter to reflective propagation.

The Jan 5 example with A_k 14 and Storm Index -25 shows a normal magnetically disturbed day with 440 cps tone amplitude fluctuating about the 600 cps tone amplitude level and a db spread of about 6.

Inversions of the twotones occur several times, and could conceivably be related to varying interference patterns caused by phase variations at the two audio wavelengths, as suggested elsewhere.

Fig. 8 is worth a brief look as illustrating the machine-like regularity the rotations can achieve, and sustain for hours, on a magnetically very quiet day.

Fig. 9 deserves a second look, showing the difference between simultaneous recordings at Concord and Scituate on 4 April 60, following the vast magnetic and ionospheric disturbance starting the end of May. A_k for the 4th was back down to 18, Storm Index to 37.

In contrast to the quiet-day close similarity of signals at the two stations shown in Fig. 1, they now appear vastly different, but still contain close corre-

lations. The Concord signal shows slightly ragged Faraday-type rotations, with large and apparently random amplitude variations, and almost no discernible trace of frequency selectivity or 'two-tone remodulation'.

The Scituate signal, on the other hand, shows as clear and regular amplitude separation of the two tones and the carrier as in Fig. 2. A novel factor here is that the separated tones at Scituate show the same interference pattern as appears on the nearly pure rotations at Concord, with very nearly the same rotation periods at a given time. The general amplitude levels at the two stations are of the same order, except for the 10 db plus depression of the carrier at Scituate, nearly to the noise level. A further significant factor here may be that the Scituate interference pattern on the carrier is slight compared to that on the two tones, although this may be merely a function of the far smaller carrier amplitude.

Fig. 10 is another illustration of the strong, but not immediate, negative correlation between the percentage of hours per day that Faraday appeared, and the Belvoir daily A_k variations for Feb 60. Storm Index figures were available for only part of the month.

The month was quiet magnetically, so the rotations never disappeared totally. On the 16th and 17th only did the A_k ever exceed 20, slightly. On these days the Ionospheric Storm Index shot up to plus 13 and plus 39, respectively. The percentage of Faraday dropped nearly to zero on the 16th, rose slightly above 30 on the 17th and 18th, then commenced a gradual descent until the 21st. Then it started a gradual climb, with A_k values remaining well under 10. The correlation between A_k and percentage Faraday for this month was -0.68. This data has not been studied for predictive possibilities.

A description of some major features of the two puzzling, allied anomalies has been given, with some evidence of correlation with magnetic variations and certain types of Sporadic E. Like the Sporadic E, no very satisfactory explanation of the phenomena has yet been found.

Refined instrumentation, notably a mobile laboratory operation and use of directive antenna and polarimeter observations, is being initiated for more refined study of the phenomena. An important objective of the study will be to attempt to explore and apply the apparent anticipatory effect of major changes in magnetoionic conditions.

ACKNOWLEDGEMENTS

We are indebted to many people for aid in attacking this puzzle. First to Dr. Philip Newman, of Propagation Sciences Laboratory, AFCRC, for interest, support and encouragement in the study.

To Mr. E. Weiwara and Mr. James Weldon, of CRPL Observatory at Ft. Belvoir, Va., and also to Messrs. F. Dickson and Harry Busch, of Signal Corps Propagation Agency, Ft. Monmouth, Va., for making available detailed sounding and other helpful data.

And in-house, to SSGT Billie R. Jones, and Mr. W. H. Cameron, for aid in operation and scaling and study of the data, and some special observations, as well as for special calculations by R. A. Swirbalus.

REFERENCES

1. Ellyett C. and Watts J. M.: "Stratification in the Lower Ionosphere" -- Jour. of Research NBS-D Radio Prop. Vol 63D, No. 2, Sept-Oct 1959.

2. Gardner F. F. and Pawsey J. L.: "Study of the Ionospheric D Region using Partial Reflections" -- J At. Terr. Phys. 3, 321, (1953).

3. Hibberd F. M.: "Ionospheric Self-Interaction of Radio Waves" -- J At. Terr. Phys. Vol 6, No. 5, May 1955, p. 268 ff.

4. Hibberd F. M.: "Self-Distortion of Radio Waves in the Ionosphere, near the Gyro Frequency" -- J At. Terr. Phys. Vol 11, No. 2, 1957, p. 102 ff.

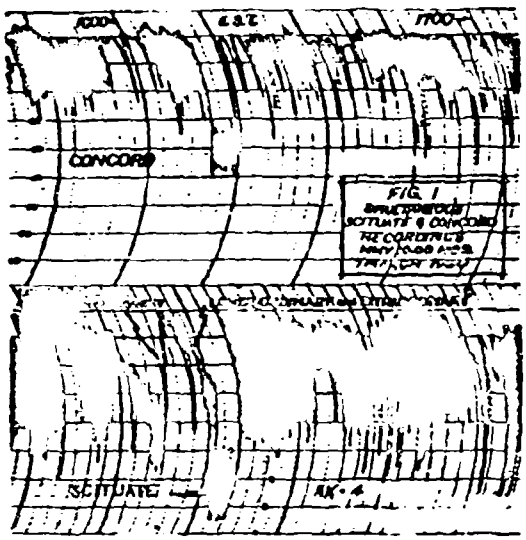
5. Bailey D. K.: "Disturbances in the Lower Ionosphere Observed at VHF" -- Jour. Geophys. Research, Vol 62, No. 3, Sept 1957, p. 431 ff.

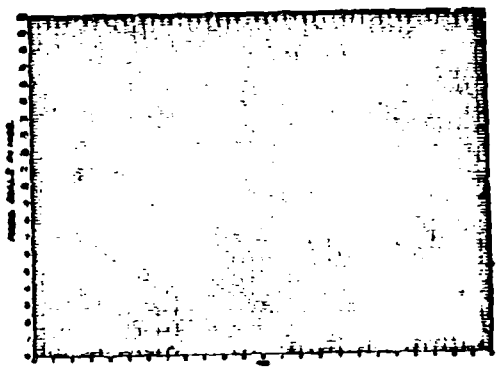
6. Nicholson P. F.: "The Nature of Fading Patterns of Oblique Incidence Radio Signals Received at Middle Latitudes During Certain Ionospheric Disturbances", -- School of EE, Cornell University, Ithaca, NY, Research Rpt EE 372 (Tech Rpt No. 47, 15 Aug 1958).

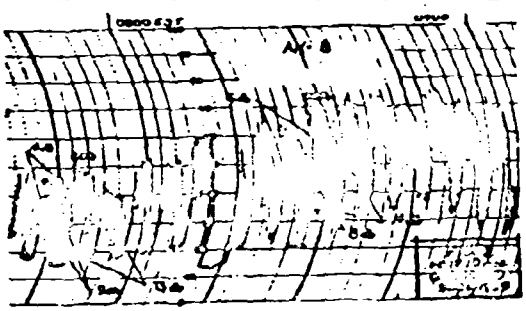
7. NBS Tech Note 2; Zacharisen D. M.: "World Maps of F 2 Critical Frequencies and Maximum Usable Frequency Factors" -- Apr 1959, PB 151361.

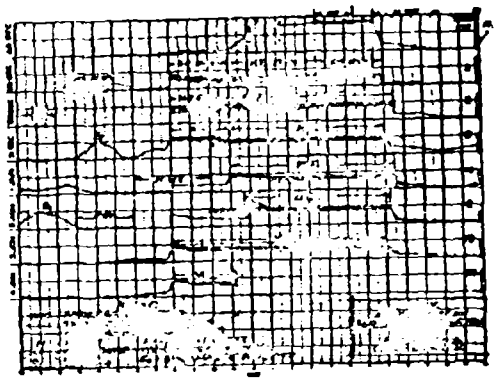
8. Marsh H. S., Cameron W. M., Swirbalus R. A.: "An Index of Ionospheric Storminess", AFCRC Tech Rpt TR-59-365, Dec 1959.

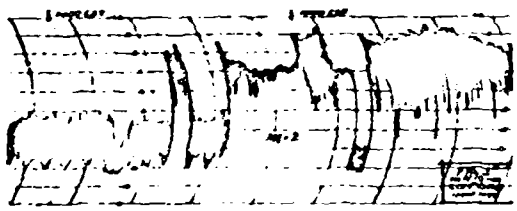
9. Layzer D. Prof.: private communication.

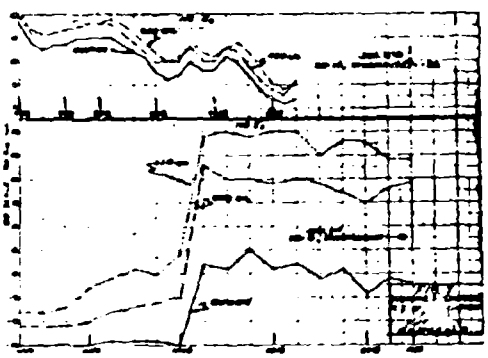

CONCORD
SCITUATE

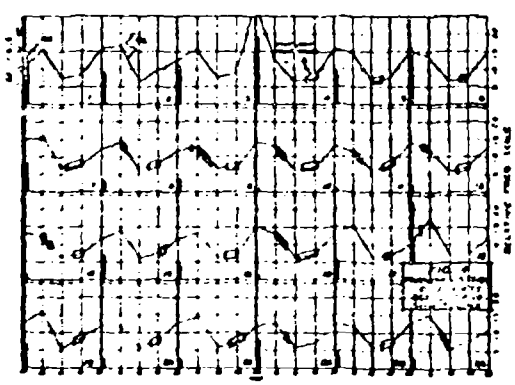


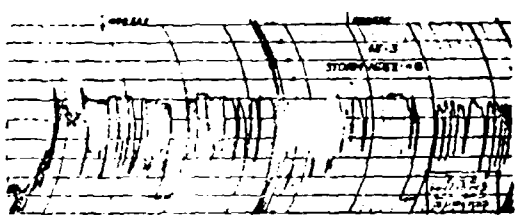


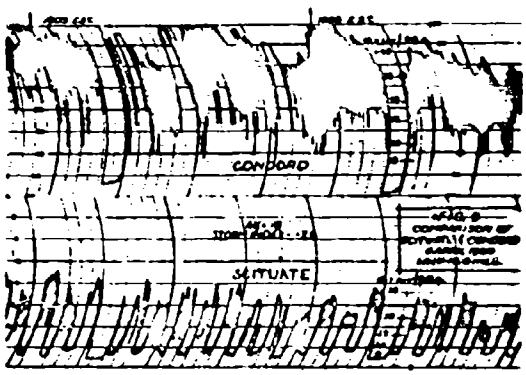


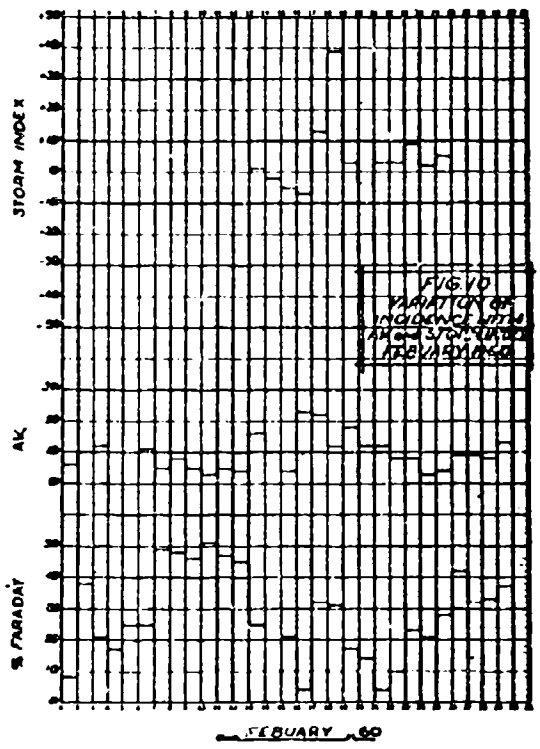









CONCORD
SITUATE


STORM INDEX
AK
% FARADAY
FEBUARY 60

DIGITAL BATTLEFIELD COMMUNICATIONS

By: Mr. W. C. Slagle, Project Engineer, Stromberg-Carlson - San Diego

Introduction

BASIC is digital communications equipment developed by Stromberg-Carlson - San Diego for the United States Marine Corps. The name BASIC stands for Battle Area Surveillance and Integrated Communications. The digital communications system is composed of an input device known as the Message Generator, a voice-transmitting Communications Link, and output equipment known as the Terminal Equipment.

In the fall of 1957, Stromberg-Carlson - San Diego engineers went on maneuvers with the Marines at Camp Pendleton to observe the need for a portable digital communications system. As a result, a proposal was made to the Marines for Digital Communications Equipment to be used on the battlefield. The task was broken into three phases: Phase I - Proof of Feasibility, Phase II - Field Evaluation Equipment, and Phase III - Operational Equipment.

Phase I - Proof of Feasibility

Phase I was completed in December, 1958, with the successful demonstration of feasibility at Camp Pendleton. The input device was a message generator employing a telephone dial. Laboratory equipment which produced an electronic display on a CHARACTRON Shaped Beam Tube served as the output device. Communications were demonstrated over wire lines and radio.

Marine Corps personnel carried five of the telephone-dial-type message generators into a simulated battle situation. The telephone dials contained, in addition to the normal 10 digits, five standard military symbols, four compass headings, and a space. The message consisted of target geographical position, target type, and general target direction of movement. A map overlay is used on the CHARACTRON Shaped Beam Tube display unit to show the reported information in its correct geographical position. Friendly and enemy forces were directed from the same room, so that observers could see the accuracy of the information reported.

Phase II - Evaluation Equipment Operation

In March, 1959, Stromberg-Carlson - San Diego began Phase II and developed the evaluation equipment described below. Essentially, it consists of the Message Generator and interpretive output equipment known as the Terminal Equipment.

The Message Generator for Phase II is a burst type device rather than the dial box used in Phase I. The information reported includes target type, target quantity, target position, target direction of movement, time of sighting, and observer identification. The transmitted message contains 22 characters. Twenty-one of the characters are derived from thumb-operated switches and one character is permanently wired for observer identification. Since the Message Generator produces an audio signal, it is connected to the microphone input of the communications equipment. Although the microphone for voice communication is connected to the Message Generator, there is no interference with normal voice operation. Voice and digital messages may be transmitted simultaneously.

The thumb-operated information switches, which have 10 positions each, are marked for an intelligence application. Eighteen symbols contained on two switches are available for describing a target. The 10 digits of position information (five for X and five for Y) allow target position reporting with an accuracy of 10 meters. A single individual or up to nine divisions may be reported.[1]

[1] Two quantity characters are provided, the first having positions 0 through 4 and F, C, B, R, and D. The second character is 0 through 9. Thus from one to 49 individuals or from one to nine platoons, companies, battalions, regiments, or divisions may be reported.

tube until they lie on the reference marks of the map holder. Map scales of 1:50,000; 1:100,000; and 1:250,000 may be used. Scale switches on the control panel adjust the electronic presentation for the map scale in use.

An electric typewriter records the target information as it arrives, providing a permanent record of all digital messages on a 3-inch x 5-inch card. The card stock is a continuous ribbon of fan-folded, pin-feed paper perforated to tear down to the 3-inch x 5-inch card size.

A punched paper tape reader associated with the typewriter permits the addition of static information such as the date to each message card. This is accomplished by switching the typewriter control to the tape reader at the completion of message recording. The desired information, contained in a paper tape loop, is then recorded. Controls which are duplicates of the erase and new target controls contained on the display portion of the terminal equipment are available at the electric typewriter. Also associated with the electric typewriter is a paper tape punch which produces a paper tape record of all information recorded by the typewriter.

Phase II - Evaluation Equipment Packaging

The Message Generator weighs approximately nine pounds and is 11 3/4 inches long, 5 inches wide, and 5 1/2 inches deep. The circuitry is contained on 11 printed circuit boards. Eight boards lie lengthwise in the case. The other three boards lie crosswise and between the rows of input switches, which are contained in three rows of seven switches each.

The Terminal Equipment is contained in five field cases. These are the display case, the audio box, the power case, the logic and storage box, and the Flexowriter case. Standard commercial Stromberg-Carlson components are used in the terminal equipment, but the packing density is greater than for commercial equipment. The cases were designed for optimum portability from a weight standpoint. The storage element, a magnetic drum, is double shock mounted and is contained in an insulated enclosure.

The display case has a leg which may be used to support the viewing end of the case while the other end may be rested on top of the audio box. This allows knee space for a seated operator. The support leg may be folded and latched to the bottom of the display case if the display is stacked on more than one of the other four cases or on a jeep or other vehicle.

Phase II - Evaluation Equipment Description

The primary purpose of the BASIC equipment is digital communication, but the present voice capability is also retained. Simultaneous digital and voice transmissions are made possible by a notch filter in the audio pass band of the voice transmission equipment. Frequencies removed from the voice spectrum are centered at 2570 cycles per second and extend from 2184 cycles per second to 2995 cycles per second. The digital information is transmitted inside this band by frequency shift keying. The frequency shift is 75 cycles per second from center; a "1" is 2645 cycles per second and a "0" is 2495 cycles per second. The information transmitted by the Message Generator consists of a fixed transmission sequence and the individual character codes. The bits constituting a message are sent serially.

The Message Generator contains a 125 cycle per second clock which is used as the timer for the Message Generator's logic. The principal components of the logic are two counters, one for sequencing through the information switches, and the other for determining the number of pulses generated for each switch setting. A series of pulses numbering from one to 10 is transmitted by each switch as it is selected, depending upon the position of the switch. The pulses passed by the switches are applied to the FSK Oscillator, then to an audio mixer. A second input to the audio mixer is the filtered voice input. The output of the audio mixer stage, which can be digital, voice, or both, is applied to the transmitter microphone input. Obviously, any audio transmission equipment can be used as the link between the basic Message Generator and Terminal Equipment.

Direction of target movement may be reported with an accuracy of one degree, and the time of sighting is accurate within one minute.

To send a message, the observer sets up the complete message by operating the input switches. He checks the entire message visually and corrects any errors that may be present. When he is satisfied that his message is correct, he pushes the transmit button. This triggers the Message Generator, causing it to send the entire message in a burst over the Communication Network. Maximum transmission time is approximately 3 1/2 seconds. The observer leaves the message set up intact until he knows that the message is received. If his message is accepted, a coded tone signal is sent over the network. Should his message not be received, the observer continues to transmit until he receives the confirming tone signal.

Each of the information switches is backlighted for night operation. These lights are controlled by a toggle switch on the panel of the Message Generator.

The Message Generator contains batteries for the operation of both the night lighting circuits and the electronic logic circuits and is completely portable. Battery life, under the normal intermittent operation, is approximately 30 days. The range of operation depends entirely on the limitations of the communications network.

The Terminal Equipment provides three types of outputs: electronic display, electric typewriter and paper tape punch.

The electronic display is derived from a CHARACTRON Shaped Beam Tube, displaying up to 50 targets at a time. All targets in the system can be viewed at once or only certain target types or categories of particular interest can be displayed. In the latter case, all other targets are blanked out without being removed from the system.

The eight basic target categories are: troops, vehicles, armored vehicles, weapons, ships, observers, electronic installations, and aircraft landing fields. In addition, an override switch is provided to permit the operator to display all eight categories simultaneously.

In addition to selecting specific categories of targets for viewing, the operator may also control the amount of information shown for each target. The target information display format is divided into four portions known as features. These features are the minimum target format consisting of two target symbols, three heading digits, two quantity symbols, and one observer identity symbol; the message number; the time of observation; and the target coordinates. The minimum target format is always displayed for each selected target type. The latter three features may be independently suppressed by the operator with the feature switches on the control panel.

The Terminal Equipment assigns the message number. This permits the operator to erase selected targets. An alarm indicator lamp signals the operator when the system contains 50 targets and can accept no more new messages.

When a new target entering the system is of a category selected for display, its format blinks, gaining the attention of the operator. As an added precaution, an alarm lamp indicates the presence of a new target in the system because the target's geographic position may be outside the viewing area in use, or may be of an unselected category. The operator, by means of a new target acknowledge control, can extinguish the light and end the format blinking.

The displayed information is optically superimposed on a standard military field map portion approximately 21 inches square. The paper map, after it is folded and attached to a map carrier frame, is manually inserted into the display enclosure. The map is accurately aligned with the electronic presentation by matching fiducial marks on the CHARACTRON Shaped Beam Tube and reference marks on the map holder. The horizontal and vertical map coordinates which correspond to the reference marks on the map holder are entered into the deflection system by digital switches on the control panel. When the map is changed, the operator moves the horizontal and vertical fiducial marks on the display

Redundant transmission is used as the method of accuracy checking. Each message is transmitted twice on a character-by-character basis. That is, the code for switch number one is transmitted twice, then the code for switch number two is transmitted twice, etc.

Pressing the transmit switch activates the "ON" relays to apply power to both the communications transmitter and the Message Generator electronics. An initial delay of approximately 272 milliseconds is built into the Message Generator to allow the transmitter to stabilize and to allow any turn-on transients in the logic to settle. The initial delay is composed of two pulses each followed by a 132 millisecond delay. After the initial delay, the logic is reset to begin the message sequence.

The switch sequence counter selects the information switches one by one in a fixed order. Each switch is dwelt upon long enough to accomplish the redundant character transmission. The digit counter and the information switch position determine the number of pulses transmitted to represent the character. Position Zero for each information switch is represented by a single pulse; Position One by a series of two pulses; Position Nine by a series of ten pulses. The Message Generator identity symbol is wired into the circuit permanently and may be thought of as a fixed information switch.

The inter-digit delay, the time between character codes, is fixed at 20 milliseconds.

The Terminal Equipment receives its input from a standard communications receiver, and the audio signal is split into the digital and voice channels. The digital transmission can be detected on the voice channel as a faint noise in the background. The voice of the person speaking is as readily recognizable either with or without the filter.

The digital information is applied to an FSK detector where it is converted to voltages representing ones and zeros. These binary voltages are applied alternately to each of two input counters. The counters have two functions: (1) To convert the serial code to four-bit parallel code, and (2) To provide comparison for the accuracy check.

The inter-digit delay produced in the Message Generator is detected and used to gate the incoming bursts of pulses alternately to the input counters. The redundant character transmission allows a comparison to be made as each character is received. If the two character transmissions compare exactly, then the four-bit word resulting is passed to a word assembly unit where it is converted to a seven-bit Flexowriter code. The seven-bit word is stored on a magnetic drum[2] for later use.

The word assembly unit originates the message received signal. This occurs only if 22 characters have been received correctly. The word assembly unit also assigns a message number to each message accepted. Messages are numbered from one through 99. When the number of received messages exceeds 99, the numbering starts again at one.

The seven-bit characters stored on the magnetic drum provide the information for both the electric typewriter and the CHARACTRON Shaped Beam Tube display.

The electric typewriter records the entire message on one line of a 3-inch x 5-inch card. A punched paper tape record of each message can be made by activating the paper tape punch associated with the typewriter.

Every message received is available for display on the CHARACTRON Shaped Beam Tube. However, every message in the system is not necessarily displayed. By use of the category and feature select switches, the console operator determines which messages are displayed and how much of each message is displayed. The category select controls compare the codes of the military symbols in each message with types of targets specified for viewing. The resultant signals control the unblank section of the display. The feature select control determines the number of format lines displayed per target. The minimum display is two lines; the maximum five lines. Since the electronic display is registered to a field map, there is no reason for

2 The magnetic drum storage is required as a recirculating storage for the electronic display.

printing out the geographic coordinates of the target as part of the minimum format. However, the maximum format does include geographic coordinates for use in laying fire.

The magnetic drum operates as a timing and control device as well as a memory device. Display tube deflection information is obtained from the reported map coordinates. The X and Y coordinates are converted to deflection voltages, taking into account the reference coordinates and the map scale.

Once the deflection voltages have settled, the format may be displayed. The initial deflection voltage is maintained constant and represents the format position. Spacing between characters in the format is produced in the Format Generator which emits incremental X and Y deflections which are added to the original format position. A format blink control receives a signal when a new target is written on the drum. The blink circuit alters the output of the unblank generator to cause the format for the just-received target to blink. An alarm light is also lit for a new target. This light is required because during display operation a new target might either lie outside the viewing area or be of an unselected category and thus go undetected, if new target blink alone were depended upon.

Magnetic drum erase circuits are required so that unnecessary information may be erased, allowing space for new targets to be recorded. The erasure is controlled by the target number. The erase controls consist of information switches similar to those used on the Message Generator and an erase button. The number of the target to erased is entered into the information switches and the erase button is pressed.

The field map and the electronic display are optically superimposed by a half silvered mirror. The map shows through the mirror, and the electronic data is reflected from the mirror. The use of a standard military paper field map eliminates the requirement for any slide making or special map preparation.

Controls for system power are placed at both the display unit and the electric typewriter. Erase controls and new target indicators are available at both places. Signals constituting a test pattern and CHARACTRON Shaped Beam Tube Alignment Controls are supplied as part of the display unit. Terminal equipment power requirements are 3000 watts at 115 volts, 60 cycles per second.

PARTICIPATION IN VARIOUS DATA EXCHANGE PROGRAMS

By

S. Pollock, U. S. Naval Ordnance Laboratory, Corona, California

INTRODUCTION

The purpose of this paper is to describe the participation by the Naval Ordnance Laboratory, Corona (NOLC), in three separate and distinct Reliability Data Exchange Programs. These are: the Fleet Ballistic Missile Weapon System (FBMWS) (POLARIS) Component Reliability History Survey (CRHS); the Interservice Data Exchange Program (IDEP), and the Bureau of Naval Weapons Guided Missile Data Exchange Program (G/M DEP).

Some of the major problems encountered in the performance of the FBMWS (POLARIS) Component Reliability History Survey are presented, and some measures that would alleviate these problems and thereby improve the usefulness and effectiveness of such a program are proposed. However, it is not intended to discuss the merits of any particular data exchange program.

OBJECTIVE

One of the major problems in achieving the expressed or desired reliability goal in a missile system lies in the selection of sufficiently reliable parts and components. The design engineer, working in a new missile program, is greatly in need of part/component information that is as complete and up-to-date as possible. Such information may be produced by a test program, running concurrently with the development of the new missile. However, the selection of parts and components is a function of time, money, quantity, and priority of effort. The problem would not be paramount if the designer had ample time to base his selection on sufficient quantities of test data derived from statistically significant quantities of parts and components. While the parts/components are in themselves costly, the testing is many times more costly. Often, because of work priorities, the testing laboratory does not have sufficient time to perform a complete or exhaustive test, and, more often, the limited quantities of parts/components furnished for testing are insufficient to provide decisive data. Therefore, the selection of a part/component, from a reliability standpoint, is often a compromise because of the limitation of reliability data regardless of the manifestation on the part of the component manufacturer as to reliability of his product.

Often the designer may use information that he personally acquired in other programs, but it is almost impossible for him to keep abreast of the latest reliability information on either old or new parts/components. Naturally, many of the younger design engineers have even less part/component information readily available to them.

DISCUSSION

FBMWS (POLARIS) Component Reliability History Survey

In the early part of 1958, NOLC was assigned a component reliability task by the Special Projects Office of the Bureau of Ordnance, Department of the Navy. The purpose of this task was to provide assistance to the prime contractor's designers in the selection of reliable parts/components for the POLARIS missile. This objective was to be accomplished by the utilization of applicable part/component reliability information generated in other missile programs. It was recognized that a large number of corporations, involved in the research and development and production phases of various missile programs, had acquired a considerable amount of data pertaining to the reliability of various parts/components. Further, it was apparent that this information rarely went further than the files or archives of the activity which generated or collected it.

It was therefore believed that both time and money could be saved in the POLARIS program if all available information could be obtained, analyzed, evaluated, summarized, and made available to the design engineer to assist him in the selection of reliable parts and components. In particular, reliability information was desired on electronic, electrical, mechanical, electromechanical, hydraulic, electro-hydraulic, and pyrotechnic parts and components.

In order to obtain as complete and valid reliability information as possible on the parts/components tentatively selected, it was decided to solicit information from many varied sources. These included the various missile manufacturers under contract to the U. S. Army, the U. S. Air Force, and the U. S. Navy; miscellaneous testing laboratories and similar facilities, both private and government; field testing or firing facilities; and vendor or part manufacturers. The type of information desired included engineering and evaluation tests performed on the part or component. Additional information was sought at the incoming-receiving inspection level and at various stages of the manufacturing process, including missile final acceptance tests.

Although the exact methods may be subject to disagreement, it is usually agreed that the use of general failure rate information is an acceptable means of predicting reliability. The prediction process consists, in part, of the application of engineering judgment in comparing known failure rates of parts/components in existing systems to predict probable failure rates of similar parts/components to be employed in new

applications or systems where the functional requirements of the part/component have a high degree of similarity. A more accurate prediction can be made if the environment of each part/component is also considered. The total amount of information desired includes:

1. Total quantity (of a particular part/component) for which data are available.

2. Number of failures (of this part/component) over a specified period of time.

3. Time to failure (operating time).

4. Mode of failure (electrical, mechanical, thermal).

5. Stress on the part/component at the time of failure (the load conditions at the time of failure, i.e., voltage, pressure, temperature, vibration, etc.).

In acquiring data for prediction purposes, every effort is made to include only the reliability information on similar parts/components experiencing similar conditions of shock, vibration, acceleration, temperature, pressure, and humidity as are specified for the FBM/S.

The data obtained in the performance of this survey are screened, evaluated and summarized. Engineering reports are prepared and are made available to the prime contractors' reliability groups for integration with contractor-generated data, analysis, and dissemination to cognizant design engineers.

It is to be emphasized that the reliability history survey program conducted by NOLC is not intended to replace any evaluation testing normally performed by the prime contractor, but is intended to augment his testing information and to identify peculiar or critical characteristics to aid in test planning and to minimize actual testing of parts and components for certification or qualification. However, the material in the reports may be used to serve as the primary basis for selecting the most reliable parts/components for the intended application. Further, reports may also be used to substantiate the initial selection of parts/components which are often chosen with limited reliability background data. They may be used to indicate or predict possible causes of failure that might be experienced, either during the research and development phase, the production phase, or the in-use or service phase. In addition, they provide a reference for steps that could be taken to minimize failures by employing the principles of derating, substitution circuits, or redundance. Without this information, reliability all too frequently becomes a post-mortem process whereby a great amount of effort and money are required to correct unsatisfactory weapon performance in tactical use. However, it is desired to stress the fact that the reports prepared by NOLC do not include any recommendation or endorsement pertaining to the use of any particular part or component by brand or trade-mark name.

In the beginning of the survey, NOLC prepared formal reliability reports for the more critical parts/components. These were intended to present a brief history of the development of the part/component being investigated, the causes of unreliability as discovered from the experience of other users, preventive measures recommended to minimize or eliminate deficiencies, and a statement covering the intended use of the part/component within the FBM/S and including the normal stresses that might be experienced by these parts/components.

At present, the majority of the reliability reports are submitted in the card form illustrated in Figure 1. This report form was developed to replace the informal or letter type of reliability report which was also employed at the beginning of the survey. Information is presented in two ways. The upper part provides for data to be placed in such a manner as to quickly identify the part or component that is being investigated by name and by digital generic code, part number, manufacturer model and type, the test activity that performed the tests, the quantity tested, for what programs, type of tests performed, the results of the tests performed, and any unusual or mitigating circumstances (such as being tested above or beyond the vendor's specifications), or any special instructions or specifications required in order to fully interpret the method of testing and results. At the lower edge of the card form, a microfilm strip is added. This contains the complete test report, or significant portions of the report, reduced at the ratio of 22-to-1. The engineer, if interested in more detailed data than briefly stated in the upper portion may then view the microfilmed data for the complete test results. By utilizing an appropriate viewer, the data on the microfilm is presented for easy reading on a large ground-glass screen. A special feature of the viewer employed at NOLC is an integral printer which enables the operator, or design engineer, to acquire immediately a letter size print of the data on the screen. This print then provides a ready reference for the design engineer to use at his convenience.

NOLC has employed this form for approximately 10 months and has found it to be extremely effective. It permits the acquisition, processing, and dissemination of data in a comparatively short period of time. Further, it permits the design engineer who ultimately receives the report to have as complete a record as possible of the tests actually performed, and eliminates the possibility that someone, in the process of screening, analyzing and evaluating the data, has either omitted important details or has permitted an interpolation or interpretation of data which may influence the results. For ready reference by engineers, many microfilmed reports can be kept in loose-leaf notebook binders; when no longer pertinent to a current study, they require a minimum of storage space.

Interservice Data Exchange Program (IDEP)

Because of the favorable reaction resulting from an informal survey conducted among various ballistic missile contractors during the latter part of 1958, and because of the encouraging results being obtained and manifested in the Air Force's TITAN reliability exchange program and the Navy's FBMWS (POLARIS) Component Reliability History Survey, representatives of the three services proposed that a program be initiated for the free interchange of parts/components data among the Department of Defense ballistic missile activities. This proposal was accepted by the three services and has resulted in the establishment of the Interservice Data Exchange Program for ballistic missiles and space systems, known as IDEP.

During calendar year 1959, and the early part of 1960, much effort was exerted by representatives of the three services including NOLC toward establishing the administrative details of this program, which included the format to be used by the participating contractors, and by the Data Distribution Centers (DDC), to transmit and receive test report information so as to result in the maximum utilization by the recipients, as well as establishing a comprehensive Part/Component Digital Generic Code acceptable to all three services.

Because of the participation of NOLC in the FBMWS CRMS Program, it has been designated one of the three Data Distribution Centers.

A few of the objectives of this program are:

1. To reduce duplicate expenditures for parts testing.

2. To avoid repetition of tests already accomplished and provide prompt indication of failure modes, thereby speeding up new projects.

3. To encourage standardization of test methods, levels, reporting, and specifying as an incidental result of widespread observation of varied techniques.

4. To lead eventually into a voluntary intercontractor preplanning of complementary test programs. The data required include, but are not limited to, test results such as:

   a. Qualification or Certification Tests

   b. Production Acceptance Tests (if particularly significant)

   c. Diagnostic or Design and Development Tests

   d. General or Comparative Evaluation Tests

   e. Reliability, Exaggerated Stress, and Life Tests

Each service has stated that reports of tests conducted on parts/components by their contractors engaged in the development and production phases of ballistic missiles and space systems shall be transmitted to the DDC specified by the cognizant service. For example, those Navy contractors engaged in the development of a ballistic missile weapon system which have been selected to participate in this program, forward their data to NOLC.

No summarizing or editing is done by the DDC. Its principal function is to reproduce and distribute the test results submitted by the contractor. Each participating contractor summarizes the results of his test on a Standard Report Summary Sheet which is forwarded, with the complete test results, to the DDC. By multilith process, the information on the Standard Report Summary Sheet is transferred to a 5x8-inch Report Summary Card; the test data itself is microfilmed at a reduction ratio of 22 to 1 and inserted in a Military "D" double-aperture EAM card (known as the Engineers' Microfilm Card) by the DDC. (See Figure 2) One EAM card provides 17 pages of microfilmed test report information plus the microfilmed Standard Report Summary Sheet. In the event the test report has more than 17 pages, additional Military "D" double-aperture EAM cards are utilized. All pages of the test report are microfilmed. The EAM card provides, after insertion of the microfilm into the two apertures, 30 key-punch spaces for storage of the following type of information: card of cards; date of the test report; vendor; index number (based on an established Part/Component Digital Generic Code and Contractor's assigned code); type of test; the missile system or space program; the card code. The cards are also color coded, by date, to facilitate purging obsolete data from the files.

Two sets of Report Summary Cards and the Military "D" double-aperture cards are transmitted by each DDC to other participating contractors and Government agencies, based upon their applicable "areas of interest." One set of cards is transmitted to the other two DDCs to establish a complete library of cards for test reports at each Center. Ten copies of each set of cards are also forwarded to ASTIA by the DDC, along with one copy of the contractor's test report. Placement of a participating contractor on a distribution list for the Engineers' Microfilm Cards and the associated Report Summary Cards is accomplished by a formal letter request from the contractor through normal channels.

BuWEPS Guided Missile Data Exchange Program

A program similar in purpose and intent to both the FBMWS CRMS and the IDEP is being established at NOLC for the Navy's guided missile programs. The significant difference in this new program is that, in addition to reliability data,

information relative to parts and components is received and exchanged among the various prime contractors engaged in the Navy guided missile development and production program, utilizing NOLC as the Data Collection and Distribution Center (DC & DC).

In June 1959, NOLC was assigned the task of developing a method and means by which design engineers might have readily available to them current information on the reliability and performance experience of certain parts and components proposed for use in BUORD (now BUWEPS) guided missiles.

To accomplish this task a feasibility study was made which took into consideration several factors: types and kinds of data desired by reliability groups and design engineers, types and kinds of data that are presently being generated within each prime contractor's establishment, method and format to be employed by each of the prime contractors participating in this program in transmitting of data, and method and format to be used by the DC & DC in transmitting data to each of the other participating contractors.

Figure 3a presents the prime contractors contacted in this study, the specific Navy missile program, the proposed flow of data, and the type of output data resulting from action by the DC & DC.

As a result of the study, it was decided that the design engineer would gain the greatest amount of information from three principal sources. These are: the specification data sheets, the application data sheets, and the results and reports of tests performed on parts and components. All this information is transmitted by each prime contractor to NOLC (the DC & DC) for processing and dissemination. As a result of this processing, the method of presentation becomes more nearly uniform and standard. It is thus possible to place each type of data in the same kind of binder at each of the participating contractor establishments. Figure 3b is a sample of the Part/Component Test Report Summary Card with microfilm. This card is similar in appearance, purpose, and intent to the Summary Report Card employed by NOLC in the FBM/S (POLARIS) Component Reliability History Survey.

STATUS OF DATA EXCHANGE PROGRAMS

FBM/S (POLARIS) Component Reliability History Survey

To date, approximately 24 prime contractors engaged in the development and production of various missile systems, covering all three services; numerous commercial testers; private and Government laboratories; and many subcontractors and vendors have been contacted in the fulfillment of this program. Approximately 16 technical memorandum type of reports have been issued which provide a complete story on all available reliability data on a single component as well as 400 informal letter reports, including the Summary Report Cards with microfilm. These cover over 200 parts/components, categorized as electrical, electronic, mechanical, pyrotechnic, etc.

IDEP

NOLC is anticipating the first interchange of data in this program in the very near future. At the time of this writing, it appears that 1 July 1960 can be tentatively established as the initial date for flow of data.

C/N BEV

All the management and administrative details have been accomplished in accordance with the requirements of the Task Assignment. It now appears that 1 July 1960 can be tentatively established as the initial date for flow of data in this program.

PROBLEMS ENCOUNTERED IN THE FBM/S CRHS

Although this particular program may be considered to be relatively successful, it could have been more so if there had been more complete documentation of part/component reliability information. While the lack of this information can be most frustrating in a survey of this nature, it appears that documentation is being improved with each passing year. At the start of this survey, it was found that the recording of reliability data and the degree of documentation on the capabilities of various parts/components varied considerably among the missile manufacturers and testing activities. In some instances, the cost of the item being purchased did not provide for such information, and, in many instances, the contract did not clearly specify or state what type of information was wanted, much less specifying how it should be recorded. Such information as engineering type approval or qualification type testing is usually available from the manufacturer or his test facility. Information is also available at the incoming-receiving inspection level, but the amount of information varies and the tests performed at this level are quite limited. As the manufacture or assembly of the missile progresses, it becomes increasingly difficult to obtain complete data on any part or component because of the loss of identity in a subsystem, system, or assembly. To isolate or identify the cause of failure, time to failure, mode of failure, and stress on the part/component at the time of failure in a complex system is most difficult. If records are not maintained, or are incomplete, the information gleaned must be carefully evaluated to minimize the possibility of erroneous conclusions. Also, great care must be exercised in the screening of data to ensure that only data which are current and recent are used. Because of rapidly changing conditions or requirements, vendors may change or modify their product to

satisfy these needs. In fairness to the vendor, the reliability data must be based on the performance of his changed or modified product. Should the quality and reliability of a vendor's product vary from a previously established level, the information, if current, will more clearly reflect the true situation. The use by many manufacturers of an "Approved Vendors List," or its equivalent, minimizes the acquisition of unreliable parts/components. It also minimizes the purchase of unknown parts/components on a "similar or equal" basis, or on the basis of low monetary cost rather than "merit" or reliability.

Available documented field or fleet usage data have been meager because of several factors which include loss of part/component identity, lack of telemetering channels, and inherent limitations imposed as a result of the nonrecoverability of an expended missile for failure diagnosis. However, NOLC has noted a marked improvement in the quality and quantity of service reliability data, which has resulted from increased awareness of the importance of feedback of pertinent field and fleet usage data to the research, development, and production agencies.

PROPOSED MEASURES TO IMPROVE DATA COLLECTED IN A PART/COMPONENT RELIABILITY SURVEY

With the strong appeal and the absolute necessity for more reliability in our missile systems, many missile contracts now include in the procurement specifications the reliability desired of the missile system and thereby indirectly establish the reliability desired of the various subsystems, assemblies, equipments, parts and components. This has resulted in the maintenance of better records on any tests that are performed. However, omission of operating time is still a major problem, but this obstacle is slowly being overcome. It has also been noted that the missile procurement specifications now more clearly set forth the requirements for the collection, analysis, and documentation of part/component reliability data, and that greater emphasis is being placed upon the acquisition of more reliable parts and components from vendors.

CONCLUSION

In view of the greatly increased costs in the development and production of advanced weapon systems, the Navy, as well as the Air Force and Army, has found it mandatory to increase the reliability of such systems. In addition, the necessity of providing the fleet with the Ballistic Missile Weapon System at the earliest possible date has prompted the Special Projects Office of the Bureau of Naval Weapons to take the initiative in profiting from the experience gained in the development of other missile systems.

Consequently, it was decided to make as much use as possible of the reliability history of parts and components that might conceivably be utilized in the POLARIS system. Therefore, the Naval Ordnance Laboratory, Corona, has been performing (since 1958) a component reliability history survey. As a result of the establishment of an integrated data collection analysis and distribution system involving many manufacturers, research, development, and manufacturing activities, testing laboratories, and firing activities, reliability information has been obtained on over 200 parts and components.

Ample evidence is available to show that the results of such a reliability survey have been used in the selection of parts and components in the FBM/S and have eliminated the necessity for extensive and costly special testing programs.

Based on the successful experience of the Fleet Ballistic Missile Weapon System Component Reliability History Survey Program, a similar program has been established and will soon be implemented for the Department of Defense Ballistic Missile and Space Systems Programs and the Guided Missile Programs of the Bureau of Naval Weapons.

Having actively participated in all these programs from the beginning, NOLC has been pleased to note the increasing interest in reliability by all commands and, what is more important, the positive steps that are being taken to improve the over-all reliability of weapon systems.

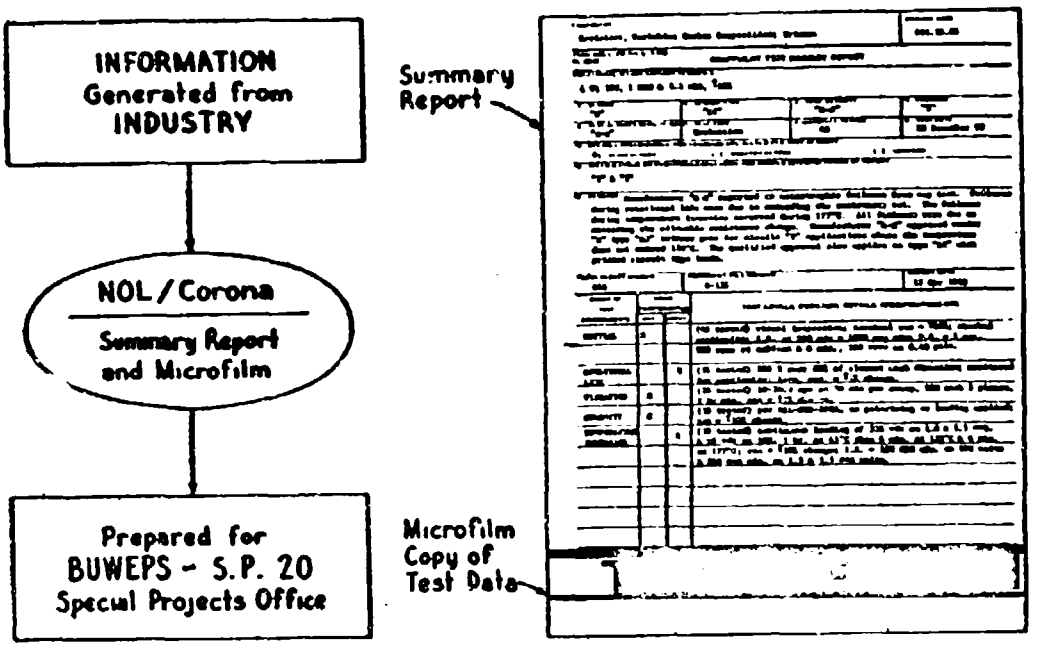


Figure No. 1

POLARIS COMPONENT RELIABILITY HISTORY SURVEY

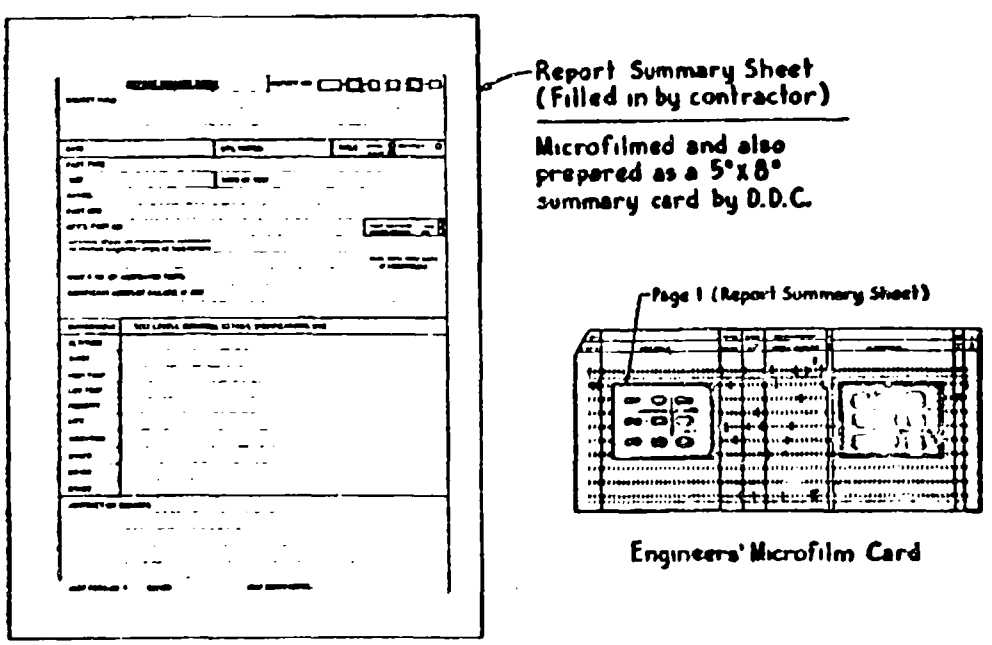


Figure No. 2

INTERSERVICE DATA EXCHANGE PROGRAM

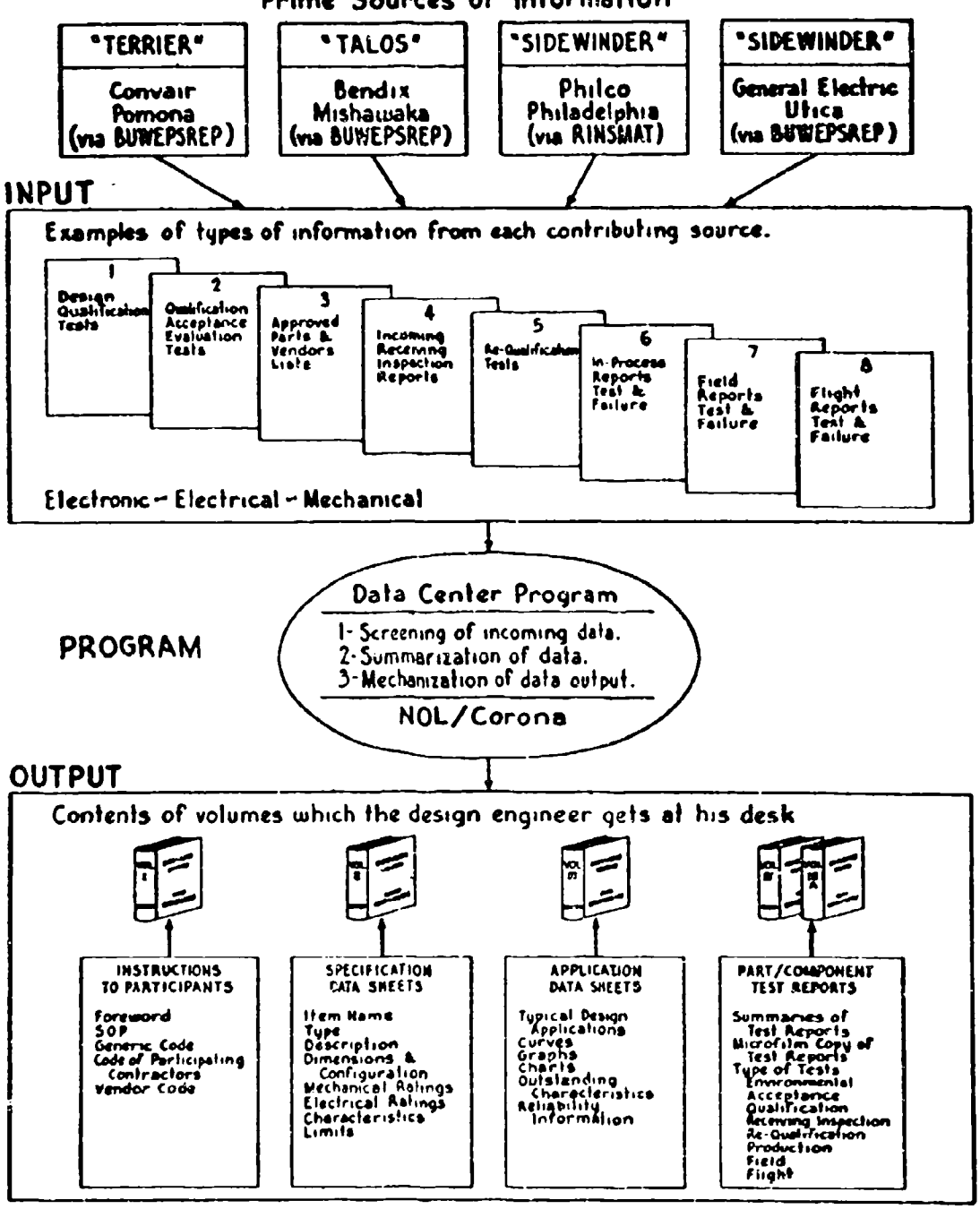
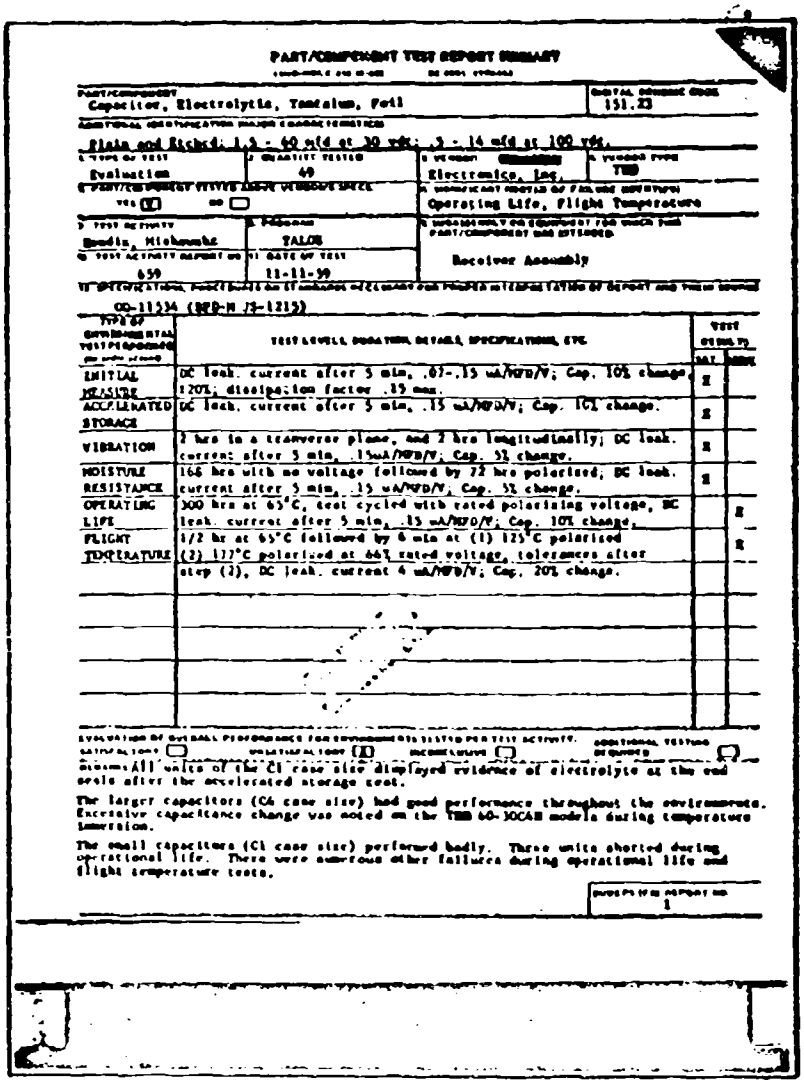


Figure No. 3a
GUIDED MISSILE DATA EXCHANGE PROGRAM



PART/COMPONENT TEST REPORT SUMMARY

Part/Component: Capacitor, Electrolytic, Tantalum, Foil — Detail Specification Sheet: 151.23

Additional Identification: Plain and Etched: 1.5 - 60 mfd at 10 vdc; .5 - 14 mfd at 100 vdc.

Type of Test	Quantity Tested	Vendor	Vendor Type
Evaluation	49	Electronics, Inc.	TBD

Part/Component Tested Above Vendor's Spec: Yes [X] No [ ]

Nomenclature and/or Part and Identification: Operating Life, Flight Temperature

Test Activity	Program	Part/Component Used On:
Goodis, Michaels	TALOS	Receiver Assembly
Test Activity Report No. 659	Date of Test 11-11-59	

Specifications, Procedures and Deviations: CO-11534 (BPO-M /5-1215)

Type of Environmental Test Performed	Test Levels, Duration, Details, Specifications, etc.	Test Results: Sat.	Unsat.
INITIAL MEASURE	DC leak. current after 5 min, .07-.15 uA/MFD/V; Cap. 10% change, 120%; dissipation factor .15 max.	X	
ACCELERATED STORAGE	DC leak. current after 5 min, .15 uA/MFD/V; Cap. 10% change.	X	
VIBRATION	2 hrs in a transverse plane, and 2 hrs longitudinally; DC leak. current after 5 min, .15uA/MFD/V; Cap. 5% change.	X	
MOISTURE RESISTANCE	166 hrs with no voltage followed by 72 hrs polarized; DC leak. current after 5 min, .15 uA/MFD/V; Cap. 5% change.	X	
OPERATING LIFE	300 hrs at 65°C, test cycled with rated polarizing voltage, DC leak. current after 5 min, .15 uA/MFD/V; Cap. 10% change.		X
FLIGHT TEMPERATURE	1/2 hr at 65°C followed by 6 min at (1) 125°C polarized (2) 177°C polarized at 66⅔ rated voltage, tolerances after step (2), DC leak. current 4 uA/MFD/V; Cap. 20% change.		X

Evaluation of Overall Performance for Environmental Test Performed per Test Activity: Satisfactory [ ] Unsatisfactory [X] Incomplete [ ] Additional Testing Required [ ]

Remarks: All units of the C1 case size displayed evidence of electrolyte at the end cycle after the accelerated storage test.

The larger capacitors (C4 case size) had good performance throughout the environments. Excessive capacitance change was noted on the TBD 60-30CAB models during temperature immersion.

The small capacitors (C1 case size) performed badly. Three units shorted during operational life. There were numerous other failures during operational life and flight temperature tests.

Sheet No. 1 of 1 Sheets

Figure No. 3b
Part/Component Test Summary Report
GUIDED MISSILE DATA EXCHANGE PROGRAM

RELIABILITY OF A PARALLEL SYSTEM CONSIDERING LOAD REDISTRIBUTION

By C. H. Tsao and H. L. Levo, Members of Technical Staff
Hughes Aircraft Company, Culver City, California

Statement of the Problem

A load is applied to a system composed of n elements (A, B, C, —). When all the elements are acting, the loads taken by the individual elements are a_0, b_0, c_0, —, respectively. When any one element has failed so that only the remaining (n-1) elements are carrying the given load, the loads taken by the individual elements are increased to a_1, b_1, c_1, --, respectively. If more elements have failed, then the loads taken by the individual elements are increased further. For element A, the probabilities of success or the reliabilities at load levels a_0, a_1, -- are A_0, A_1, -- respectively. These load levels and their corresponding reliabilities are given in Table I. Note that $A_0 \geq A_1 \geq A_2 \geq A_3$ --, $B_0 \geq B_1 \geq B_2 \geq B_3$, --, etc. Required the reliability of the system.

Solution of the Problem

Probabilities of failure of exactly one, two or three elements in the system will be analyzed in Part I. These results are then generalized to any number of element failures by the method of induction in Part II. Illustrative examples are given in Part III.

Part I - Probabilities of failure of small number of elements.

Case (1) - None of the elements fail.

Probability of success of element A = A_0
Probability of success of element B = B_0
Probability of success of element C = C_0
etc.

Probability of occurrence of Case (1) = $A_0B_0C_0$—

For the special case where $A_0 = B_0 = C_0 = $ ---, then the above expression simplifies to A_0^n.

Case (2) - One and only one element fails.

(i) Special Case - Element B is the element which fails. If the order of examination is A, B, C, --, then for the first round of examination, the probabilities for A, C, D, to succeed and B to fail are A_0, C_0, D_0, and $1-B_0$ respectively. These are listed in Table II where a circle denotes success and a cross denotes failure. Now element A must be examined again since a change of condition has occurred after the first examination; namely, element B has failed. Since element A has succeeded on the first examination, it is a certainty that its failure load exceeds a_0. After B has failed the load on element A becomes a_1 (here $a_1 \geq a_0$). Among the A elements having failure loads exceeding a_0, the fraction which have failure loads exceeding a_1 is

$$\frac{A_1}{A_0}.$$

Thus the probability that element A will pass both rounds of examination equals the probability that it succeeds on the first round times the probability that it succeeds on the second round given that it succeeded on the first round; that is, the desired probability equals

$$A_0 \times \frac{A_1}{A_0} = A_1.$$

and is listed in the last row of Table II.

The probability of having element B fail alone is therefore the product of all the terms in the last row of Table II, and is given by $A_1(1-B_0)\,C_1D_1B_1$---. For any other order of examination, such as B A B C B —, the same result will be obtained. Hence, the probability of element B failing alone is independent of the order of examination.

(ii) General Case - Any one element failing alone.

The probability of any one element failing alone is obviously given by

$$(1-A_0)\,B_1C_1D_1\text{—} + A_1(1-B_0)\,C_1D_1\text{--} +$$

$$A_1B_1\,(1-C_0)\,D_1\text{---} + \text{—}.$$

This probability is independent of the order of examination. For the special case where $A_0 = B_0 = C_0 = $ ---, $A_1 = B_1 = C_1$—, then the above expression simplifies to

$$n(1-A_0)\,A_1^{n-1}$$

Case (3) - Two and only two elements fail.

(i) First Special Case - Elements B and D are the elements which fail with B failing first.

If the order of examination is A, B, C, D, ---, then the results of examination and the corresponding probabilities are given in Table III. The probability of occurrence of this special case is $A_2(1-B_0)C_2(1-D_1)B_2$—.

(ii) Second Special Case - Elements B and D fail with D failing first.

If the order of examination remains A, B, C, D, then the results of examination and the corresponding probabilities are given in Table IV. The probability of occurrence of this special case is $A_2(B_0-B_1)C_2(1-D_0)B_2$—.

(iii) Third Special Case - Elements B and D fail in any order.

If the order of examination is A, B, C, D, --, then the probability of occurrence of this special case is the sum of the two previous special cases and is given by

$$(1-B_0)(1-D_1) + (B_0-B_1)(1-D_0)\;A_2C_2B_2\text{---}.$$

Note that this expression is invariant under an exchange between B and D. It is independent of the order of examination.

(iv) General Case - Any two elements fail in any order.

If the order of examination is A, B, C, D, ---, then the probability of any two elements failing in any order is obviously

$$(1-A_0)(1-B_1) + (A_0-A_1)(1-B_0)\ C_2D_2E_2\text{---}$$

$$+ (1-A_0)(1-C_1) + (A_0-A_1)(1-C_0)\ B_2D_2E_2\text{---}$$

$$+ (1-A_0)(1-D_1) + (A_0-A_1)(1-D_0)\ B_2C_2E_2\text{---}$$

$$+ - - - -$$

$$+ (1-B_0)(1-C_1) + (B_0-B_1)(1-C_0)\ A_2D_2E_2\text{---}$$

$$+ (1-B_0)(1-D_1) + (B_0-B_1)(1-D_0)\ A_2C_2E_2\text{---}$$

$$+ - - - -$$

$$+ (1-C_0)(1-D_1) + (C_0-C_1)(1-D_0)\ A_2B_2E_2\text{---}$$

This probability is invariant under an exchange between any two elements so that it is independent of the order of examination.

For the special case where $A_0 = B_0 = C_0 = \text{--}$, $A_1 = B_1 = C_1 = \text{---}$, $A_2 = B_2 = C_2 = \text{--}$, then the above expression simplifies to

$$\frac{n(n-1)}{2}\ (1-A_0)(1 + A_0 - 2A_1)\ A_2^{n-2}$$

Case (4) - Three and only three elements fail.

Conclusions similar to Case (3) can be obtained for Case (4). For the special case where $A_0 = B_0 = C_0 = \text{---}$, $A_1 = B_1 = C_1 = \text{---}$, $A_2 = B_2 = C_2 = \text{---}$, the probability of any three elements failing in any order is

$$\frac{n(n-1)(n-2)}{1\cdot2\cdot3}\ (1-A_0)\ [(1-A_1)(1+A_0+A_1-3A_2)$$

$$+ (A_0-A_1)\ (A_0+2A_1-3A_2)]\ A_3^{n-3}\ .$$

Part II - Probabilities of failure of any number of elements.

Let the analysis here be restricted to the special case $A_0 = B_0 = C_0 = \text{--}$, $A_1 = B_1 = C_1 = \text{--}$, $A_2 = B_2 = C_2 = \text{--}$, $A_3 = B_3 = C_3 = \text{—}$, etc.

For this special case, the probabilities of failure of precisely one, two or three elements in a structure composed of n elements in parallel are already obtained in Part I. These results are listed in Table V. By assigning values of 1, 2, 3, 4 and 5 to n, these probabilities take on special values listed in the fifth through seventh column of the same table.

It is obvious that f_1, f_2, f_3, —, are functions of A_0, A_1, A_2, —, and not of n. To determine f_1, take the case n = 1, since the sum of the probabilities of no elements failing and one element failing must be unity, we have,

$$A_0 + f_1 = 1$$

$$\text{or} \qquad f_1 = 1-A_0 \qquad (1)$$

By taking n = 2 and summing up all the probabilities under that column, we have

$$A_0^2 + 2A_1f_1 + f_2 = 1$$

$$f_2 = 1-A_0^2-2A_1f_1 \qquad (2)$$

Similarly we have

$$f_3 = 1-A_0^3 - 3A_1^2f_1 - 3A_2f_2$$

$$f_4 = 1-A_0^4-4A_1^3f_1 - 6A_2^2f_2 - 4A_3f_3$$

$$f_5 = 1-A_0^5-5A_1^4f_1-10A_2^3f_2- 10A_3^2f_3- 5A_4f_4$$

$$f_n = 1-A_0^n - nA_1^{n-1}f_1 - \frac{n(n-1)}{2!}A_2^{n-2}f_2$$

$$- \frac{n(n-1)(n-2)}{3!}A_3^{n-3}f_3 \text{ ————}$$

$$- \frac{n(n-1)(n-2)\cdots4\cdot3}{(n-2)!}A_{n-2}^2f_{n-2}$$

$$- \frac{n(n-1)(n-2)\cdots3\cdot2}{(n-1)!}A_{n-1}f_{n-1} \qquad (3)$$

The above expressions for f_1, f_2 and f_3 agree exactly with the results of Part I.

(i) First Special Case - Probabilities of failure of the elements are constant and not dependent on the number of elements acting. This applies to the common case where each element does not take on additional loads when other elements fail.

Here $A_0 = A_1 = A_2 = \text{—} = A$

$$f_1 = 1-A$$

$$f_2 = (1-A)^2$$

$$f_3 = (1-A)^3$$

$$\vdots$$

$$f_n = (1-A)^n \qquad (4)$$

(ii) Second Special Case - System fails when any one element fails. This applies to a system composed of n elements in series.

Here $A_1 = A_2 = A_3 = \ldots$

$$f_1 = 1-A_0$$
$$f_2 = 1-A_0^2$$
$$f_3 = 1-A_0^3$$
$$f_4 = 1-A_0^4$$
$$f_5 = 1-A_0^5 \qquad (5)$$

Part III - Illustrative Examples

Example (1) - In the parallel operation of voltage regulating tubes, Figure (1), the load level in each tube is a_0 when all four tubes are acting. After one tube has failed so that only three tubes are acting, the load level in each tube is increased to a_1. Similarly after two and three tubes have failed, the load level in each tube is increased to a_2 and a_3 respectively. All the tubes are from the same population with reliabilities A_0, A_1, A_2 and A_3 corresponding to load levels a_0, a_1, a_2, and a_3 respectively. Here $A_0 \geq A_1 \geq A_2 \geq A_3$. Required the reliability of the network.

Solution (i) - Assume the system to fail only when the circuit is broken. This occurs only when all four tubes fail. Then from Table V, for

$n = m = 4$, we have

Probability of failure of system = f_4

Reliability of system = $1-f_4$

The value of f_4 can be readily obtained from formulas (1) through (3).

Solution (ii) - Assume the function of the system to be sufficiently impaired and failure declared when three or more tubes have failed. Then from Table V, for $n = 4$; $m = 3,4$, we have

Probability of failure of system = $4A_3f_3 + f_4$

Reliability of system = $1-(4A_3f_3+f_4)$

$$= A_0^4 + 4A_1^3f_1 - 6A_2^2f_2$$

The values of f_1 through f_4 can be readily obtained from formulas (1), (2) and (3).

Example (2) - A weight W is supported by a structure composed of five wires connected in parallel, Figure (2). All five wires are from the same population with reliabilities A_0, A_1, A_2, A_3, and A_4 corresponding to the stress levels a_0, a_1, a_2, a_3, and a_4 respectively. The latter are the stress levels when 5, 4, 3, 2 and 1 wire respectively are carrying the weight W. Required the reliability of the structure.

Solution - Assume the structure to fail only when all five wires fail. Then from Table V, for $n = m = 5$, we have

Probability of failure of the structure = f_5

The value of f_5 can readily be obtained from formulas (1) through (3). The reliability of the structure is $1-f_5$.

If $A_0 = 0.9$, $A_1 = 0.8$, $A_2 = 0.7$, $A_3 = 0.6$ and $A_4 = 0.5$, then $f_5 = 0.01296$. If the usual formula for parallel systems with no load redistribution is used to calculate the probability of failure of the above system, then $A_0 = A_1 = A_2 = A_3 = A_4 = 0.9$, and $f_5 = 0.00001$. Hence the usual formula neglecting load redistribution predicts a probability of failure which is 99.92 percent too small.

Number of elements acting during the current examination		n	n-1	n-2	—	2	1
Number of elements already failed before the current examination		0	1	2	--	n-2	n-1
Load levels in element	A	a_0	a_1	a_2	--	a_{n-2}	a_{n-1}
	B	b_0	b_1	b_2	--	b_{n-2}	b_{n-1}
	C	c_0	c_1	c_2	--	c_{n-2}	c_{n-1}
Corresponding Reliabilities for Element	A	A_0	A_1	A_2	--	A_{n-2}	A_{n-1}
	B	B_0	B_1	B_2	--	B_{n-2}	B_{n-1}
	C	C_0	C_1	C_2	--	C_{n-2}	C_{n-1}
$A_0 \geq A_1 \geq A_2 \geq A_3$ ---							
$B_0 \geq B_1 \geq B_2 \geq B_3$ ---							
etc.							

TABLE I.

Load Levels and Corresponding Reliabilities of the Elements in a System Resisting a Given Load

ELEMENT		A	B	C	D	E	--
Results of Examinations	First Round	0	X	0	0	0	--
	Second Round	0	-	-	-	-	--
Probability of Getting the Above Results for	First Round	A_0	$1-B_0$	C_1	D_1	E_1	--
	Second Round	$\frac{A_1}{A_0}$	-	-	-	-	--
	Both Rounds	A_1	$1-B_0$	C_1	D_1	E_1	--

TABLE II.

Results of Examinations and Probabilities for the Case of Element B Failing Alone

ELEMENTS		A	B	C	D	E	--
Results of Examinations	First Round	0	X	0	X	0	--
	Second Round	0	-	0	-	-	—
Probabilities of Getting the Above Results for	First Round	A_0	$1-B_0$	C_1	$1-D_1$	E_2	—
	Second Round	$\frac{A_2}{A_0}$	---	$\frac{C_2}{C_1}$	---	—	—
	Both Rounds	A_2	$1-B_0$	C_2	$1-D_1$	E_2	--

TABLE III.

Results of Examinations and Probabilities for the Case of Elements B and D Failing With B Failing First.

ELEMENTS		A	B	C	D	E	--
Results of Examination	First Round	0	0	0	X	0	--
	Second Round	0	X	0	-	0	—
	Third Round	0	-	-	-	-	--
Probabilities of Getting the Above Results for	First Round	A_0	B_0	C_0	$1-D_0$	E_1	--
	Second Round	$\frac{A_1}{A_0}$	$\frac{B_0-B_1}{B_0}$	$\frac{C_2}{C_0}$	---	$\frac{E_2}{E_1}$	--
	Third Round	$\frac{A_2}{A_1}$	--	--	---	--	--
	All Three Rounds	A_2	B_0-B_1	C_2	$1-D_0$	E_2	--

TABLE IV.

Results of Examinations and Probabilities for the Case of Elements B and D Failing with D Failing First.

Exact Number of Elements Failing	Values of n						
	n	1	2	3	4	5	n
0	A_0^n	A_0	A_0^2	A_0^3	A_0^4	A_0^5	A_0^n
1	$nA_1^{n-1}f_1$	f_1	$2A_1f_1$	$3A_1^2f_1$	$4A_1^3f_1$	$5A_1^4f_1$	$nA_1^{n-1}f_1$
2	$\frac{n(n-1)}{2!}A_2^{n-2}f_2$	--	f_2	$3A_2f_2$	$6A_2^2f_2$	$10A_2^3f_2$	$\frac{n(n-1)}{2!}A_2^{n-2}f_2$
3	$\frac{n(n-1)(n-2)}{3!}A_3^{n-3}f_3$	--	--	f_3	$4A_3f_3$	$10A_3^2f_3$	$\frac{n(n-1)(n-2)}{3!}A_3^{n-3}f_3$
4	$\frac{n(n-1)(n-2)(n-3)}{4!}A_4^{n-4}f_4$	--	--	--	f_4	$5A_4f_4$	$\frac{n(n-1)(n-2)(n-3)}{4!}A_4^{n-4}f_4$
n	$\frac{n(n-1)(n-2)(n-3)\cdots(n-n+1)}{n!}A_n^{n-n}f_n$	--	--	--	--	--	f_n

TABLE V.

Probabilities of a Specific Number of Elements Failing in a System Composed of n Elements in Parallel. All Elements Are From the Same Population.

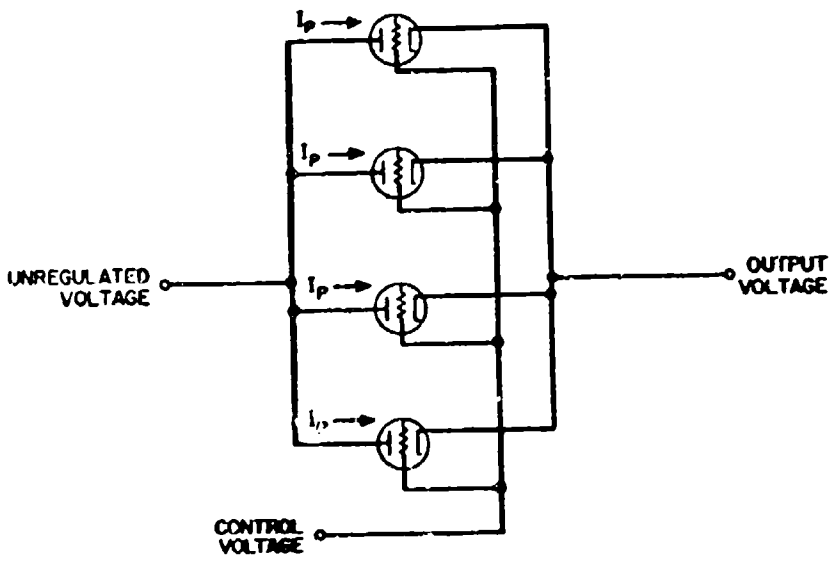


FIGURE 1. PARALLEL OPERATION OF VOLTAGE REGULATING TUBES

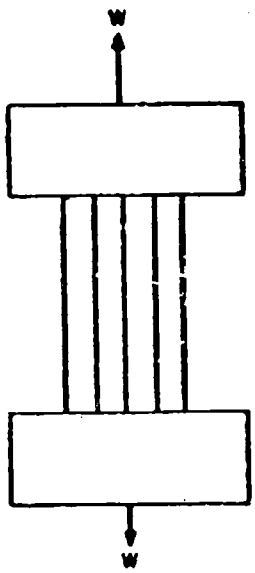


FIGURE 2. A WEIGHT W CARRIED BY FIVE WIRES CONNECTED IN PARALLEL

UNDERSTANDING AND IMPROVING SYSTEM RELIABILITY AND MAINTAINABILITY USING INFORMATION IN ENGINEERING/ENVIRONMENTAL MALFUNCTION DATA SAMPLES

By: Mr. George Allen and Mr. Albert Sutherby
Raytheon Company, Airborne Equipment Operations, Sudbury, Massachusetts

SUMMARY

This paper describes some of the activities of a Reliability Section during an airborne radar system engineering/environmental test program. These activities range from data collection, processing, and computational work to the identification of actual and potential reliability weak-links.

GENERAL DISCUSSION

One of the important provisions for a successful reliability program is the operation of system models in simulated mission environments prior to and during full scale production of systems.

While this type of testing allows reliability the opportunity to collect, process, and tabulate malfunction data, it also allows the identification of system weak-links prior to or at least simultaneous with operational or field malfunction experience, and the obtainment of corrective action which can be reflected earlier in production models.

Of course, the efficiency of a particular fix can be determined by processing appropriate field or operational information. In general, the feedback loop to systems engineering, quality control, purchasing, component engineering, and other agencies involved in the corrective action process is shortened. Timeliness of information is an important characteristic of a successful reliability program.

What type of information should reliability be interested in obtaining during an engineering test program? The basic data needed for computing system reliability is system operating time-to-malfunction. In studying a system or an ensemble of systems, the data collection process cannot begin until agreement is reached on what is meant by system operating time and system malfunction.

For the airborne radar system currently being studied by Raytheon, operating time is defined as system usage time. This system quantity is composed of transmit and standby time.

System malfunctions are divided into two general categories: lethal and nonlethal. A lethal event renders the system inoperative or dead and, therefore, completely useless to the system operator. Certainly, the operational worth of a system is partially judged by how susceptible the system is to lethal malfunctions or how successful it is in remaining on the air. Therefore, the system statistic operating time-to-lethal-malfunctions in evaluating system operational worth is considered.

One effect of system malfunctions is system down-time. The length of time that a system is down or unavailable for tactical usage, following a malfunction, is a function of

a. The severity of the malfunction, that is, the percentage of the system affected by the malfunction

b. The accessibility of the affected areas

c. Logistical problems, such as the availability of component part spares, repair channels, and repair crews

d. The availability of up-to-date check-out procedures.

Within the engineering/environmental test program, we have been concerned with two statistics: the time-to-locate-the-system-malfunction area(s) (once a search begins), and the time-to-repair, conditioned on the accomplishment of malfunction location and the availability of spares.

Maintainability numbers for a system depend on the latter statistic. Since engineering personnel were involved in the repair actions studied, the time-to-repair statistics are probably optimistic as to what can be expected during actual field operating conditions.

However, with the increasing concern for system down-time and maintainability numbers as aids in stating over-all system operational worth, it appeared that it was essential to process time-to-locate and time-to-repair data. Furthermore, when engineering personnel began to understand numerically the problems facing field personnel when using the system in the manner and for the purpose intended, it was easier to obtain engineering actions which improved field check-out procedures and made certain system areas more accessible for repair.

Component part and electron tube removals were investigated for a total operational time-in-socket value, and a correlation between removed part analysis results and reported removal cause. While computations were performed for removal rates (hours^{-1}) for a class of components, such as a particular type of electron tube, removed part analysis results were used to

a. Support engineering investigations of system weak-links of a design nature, such as socket dependency and overstressed part applications

b. Bring part vendor attention to part fabrication defects of a quality nature

c. Improve piece part incoming inspection procedures through an understanding of part malfunction modes during system operation

d. Contribute directly to the over-all improvement of field check-out procedures.

DATA COLLECTION, PROCESSING, AND COMPUTATIONAL PROCEDURES

While timeliness of information is an important consideration in planning a reliability program, data accuracy is also of concern. To help assure accuracy of information, personnel (reliability monitors) were stationed at the system test locations. The prime responsibility of these people was data collection. However, they were also able to observe and to advise Reliability whether or not the system was being used in the manner and for the purpose intended.

The usual type of data collection forms were supplied to these monitors. Each system malfunction was verified by representatives of three sections: reliability, system test, and design or systems engineering.

All malfunction data, including removed part analysis results, were encoded for IBM 650 processing. This enabled a rapid analysis of a functional block's malfunction history. Since all field malfunction data is being encoded for IBM processing, it is easy to perform comparative studies of a functional block's field and engineering/environmental malfunctions.

Statistically, the reliability function, that is, the probability of no system malfunction for a desired period of system operating time, as a function of system operating time, is expressed as

$$h(t) \triangleq \int_{t}^{\infty} f(x)\,dx, \quad f(x)>0, \quad 0 \le x \le \infty$$

where x is a continuous stochastic variable representing the length of desired periods of satisfactory operation (absence of malfunction). Desired values of x are dependent on the tactical requirements of the system. A univariate density function, f(x), satisfies the equation:

$$\int_{0}^{\infty} f(x)\,dx = 1$$

If the system property in question is operating time-to-lethal-malfunctions, the probability of no lethal malfunction, for a given period of operating time, as a function of operating time, is defined as the lethality probability function. Such a function can be written as

$$R_L(t) \triangleq \int_{t}^{\infty} f_L(\gamma)d\gamma, \quad f_L(\gamma)>0, \quad 0 \le \gamma \le \infty$$

Statistically, the maintainability function is the probability that the system will be repaired (once a repair action begins) to an in-specification operating condition within a given period of time, as a function of time. Such a function can be written as

$$m(t) \triangleq \int_{0}^{t} g(w)\,dw, \quad g(w)>0 \le w \le \infty$$

where

g(w) is the time-to-repair density;

$$\int_{0}^{\infty} g(w)\,dw = 1$$

Both the continuous maintainability and the reliability functions require a knowledge of continuous time density functions. The simplest form of such a function is the univariate uniparameter exponential density

$$Z(v) = \theta^{-1} \exp.(-v/\theta), \quad \theta>0, \quad v \ge 0$$

For a particular problem, it may be possible to construct a meaningful point type or nonparametric reliability function. Such a function can give, by observation, some indication of the underlying continuous time density function. Tests for significance can be introduced.

As mentioned before, down-time includes both time-to-repair and time-to-locate. As such, down-time is expressed in physical units (hours), whereas, maintainability is expressed in terms of a statistical quantity, namely, probability. System availability is also expressed in terms of probability and can be computed by dividing the mean-time-to-malfunction value by the sum of the mean-down-time and mean-time-to-malfunction values.

It is possible, of course, to proceed with a functional block analysis for maintainability and availability as well as for reliability. Weak-links for all three system statistics can be established by an engineering and statistical analysis of the collected information.

COMPUTATIONAL APPLICATIONS

System Life History Statistical Model

Figure 1 indicates a portion of the statistical life history model for the first system studied. The model plots the system's hazard rate (hours^{-1}) as a function of the system's operating time. Notice that the model follows the classical "bathtub" curve referred to in several reliability texts. Testing is continuing on this system and it is of interest to determine whether the increase in hazard rate (at the end of the curve) is an indication of entrance into a wear-out region which, perhaps, may be described by a Gaussian density.

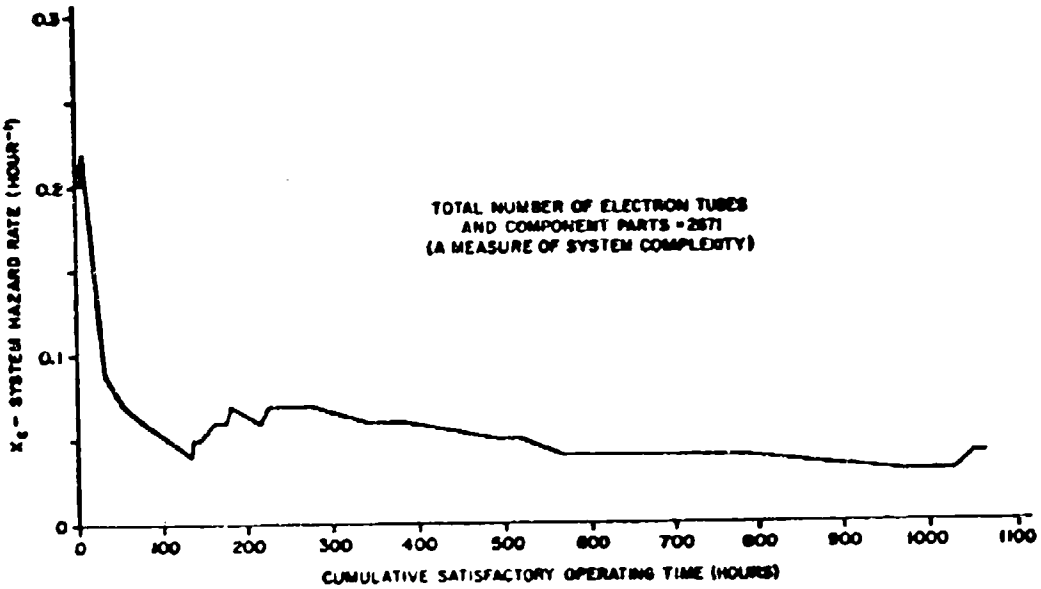


Figure 1. Airborne Radar System No. X Life History Statistical Model

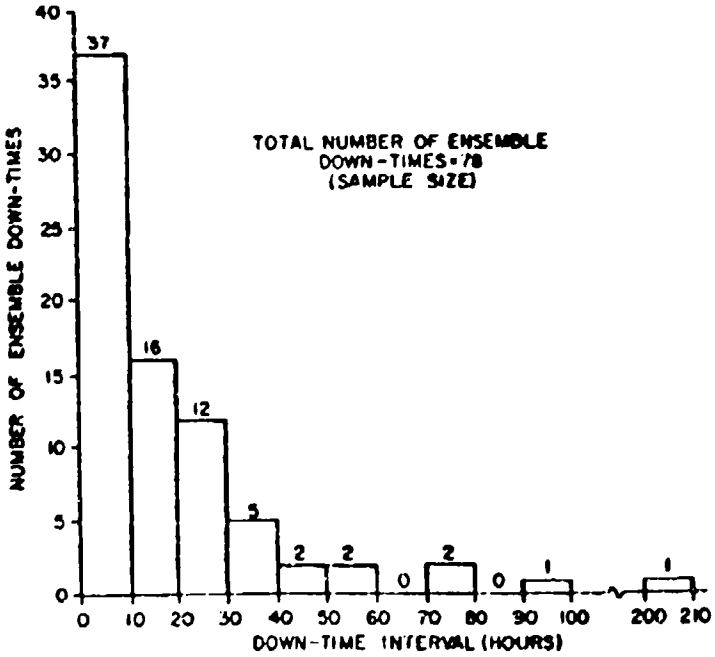


Figure 2. Down-Time Distribution for Four Airborne Radar Systems During Engineering/ Environmental Test

Down-Time Distribution

Figure 2 indicates a sample down-time distribution. This distribution was developed by pooling down-time data samples from four (engineering) identical airborne radars. The ensemble sample indicates a large number of events, 37/78 or 52.8 per cent, of short duration, that is, less than 10 hours. One event involved the 200-210 hour interval. However, 70/78 or 89.7 per cent of the events were less than 40 hours.

Figure 3 indicates two nonparametric and two parametric down-time functions. Both system down-time functions were computed using the information in figure 2. These functions describe the statistical probability that the waiting time to terminate a down-state exceeds a down-time interval, as a function of down-time. In other words, using the down-time function exp. $(-t/19.4)$, $t \geq 0$, there is a 37 per cent probability that the system will still be in a down-state after 19.4 hours.

Nonparametric and Parametric Subsystem Reliability Functions

Figure 4 indicates the present nonparametric reliability function for black box No. 1. This function was computed using both lethal and nonlethal malfunction time data developed during engineering/environmental test. The parametric reliability function exp. $(-t/63.1)$, $t \geq 0$, is presented.

This black box is a major subsystem of the airborne radar under investigation. Malfunction data from four identical (engineering) subsystems was used to develop figure 4. These four identical subsystems represent an ensemble.

Figure 5 indicates the present nonparametric reliability function for black box No. 2. The parametric reliability function exp. $(-t/73.9)$, $t \geq 0$, is also presented.

Our investigations have shown that, after suitable burn-in, the airborne radar system can be represented fairly accurately by the reliability equation

$$R_S \triangleq \prod_{j=1}^{n} R_j$$

where

$$R_j = \exp.\ (-t/\hat{\theta}_j),\ \hat{\theta}_j > 0$$

For $t > 0$, $0 < R_j < 1$. For $t = 0$, $R_j = 1$.

$\hat{\theta}_j$ represents the j^{th} subsystem computed mean-time-to-malfunction value.

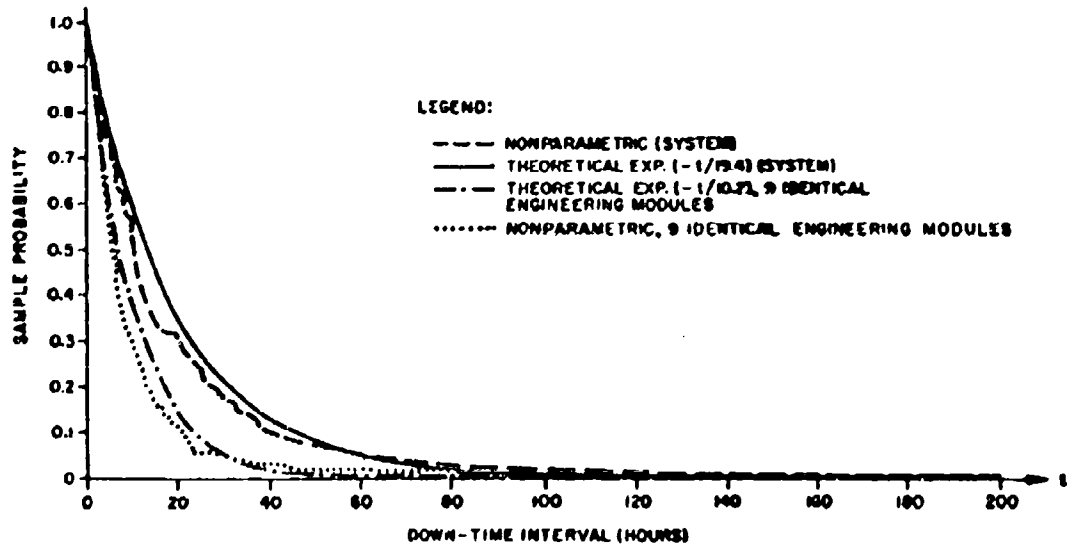


Figure 3. Parametric and Nonparametric Down-Time Functions for Four Airborne Radars Undergoing Engineering/Environmental Test

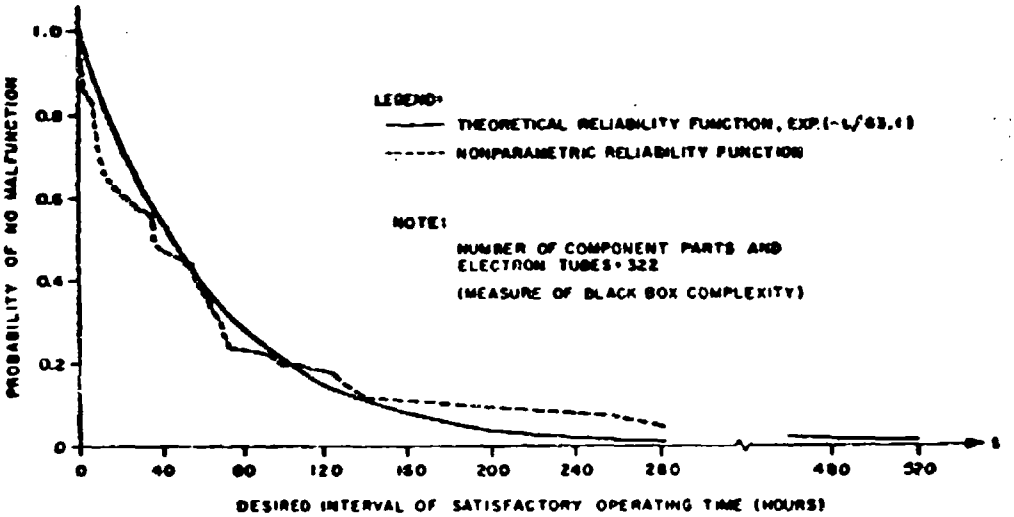


Figure 4. Parametric and Nonparametric Reliability Functions for Black Box. No. 1 Ensemble

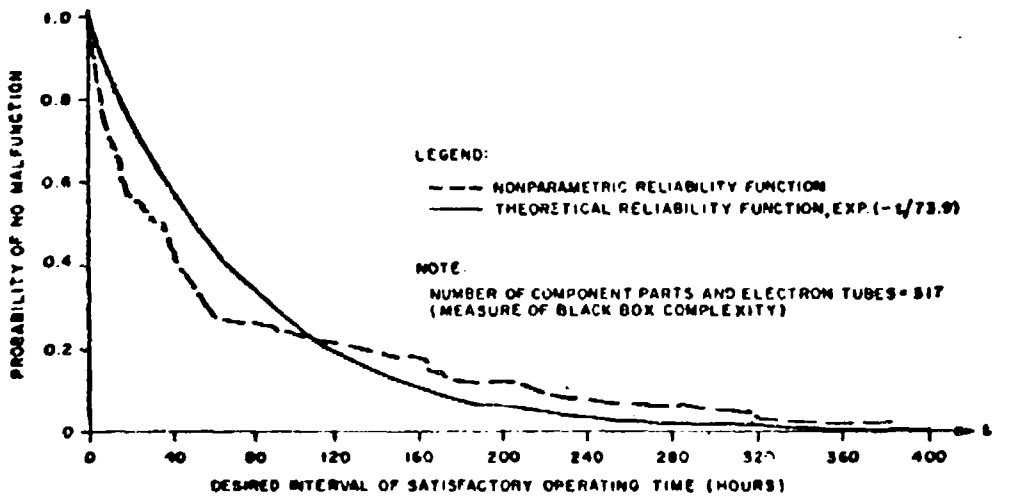


Figure 5. Parametric and Nonparametric Reliability Functions for Black Box. No. 2 Ensemble

Distribution of Component Part Operational Time-In-Socket Values

Figure 6 indicates a distribution of component part operational time-in-socket values. This distribution was developed by collecting and processing component removal information from the airborne radar system referred to in figure 1. Operational time-in-socket data was correlated with failed part analysis results and design application data in order to arrive at a family of system weak-links, recommendations for vendor changes, and recommendations for improved piece part incoming inspection tests and procedures.

Malfunction Analysis

Each malfunction that occurred during the engineering/environmental test program was classified by Reliability as to

a. Cause of malfunctional design deficiency, workmanship, test error, etc.

b. Effect of malfunction on system operation (lethal, no effect, etc.)

c. Organization responsible for corrective action (engineering, quality assurance, etc.).

Each corrective action is programmed for "break-in" into production systems. If the corrective action takes place as a result of an engineering change proposal (ECP), Reliability computes the system's expected probability or survival increase as the result of the proposed ECP action. In other words, the reliability worth of the ECP action is presented numerically to the customer.

In establishing a reliability weak-link list using the information in the malfunction data samples, repetitive malfunctions of the same type or system location are not the only criteria for a weak-link item. Lethal malfunctions are always investigated. If, for example, a relay application has malfunctioned only once but, when it did malfunction, the system was rendered unusable, the application in question would be placed on the weak-link list and corrective action would be taken to prevent the event from reoccurring.

Actually, Reliability is interested in two major types or categories of reliability weak-links: potential and actual. Actual weak-links are determined from sheer density of malfunctions in the same system location, same application, etc., or by the effect that the malfunction produces on the operational capabilities of the system. Potential weak-links are discovered, for example, by paper studies of component part and electron tube applications. An overstressed part application has the potential for producing an operational malfunction that could interrupt the operation of the system at a time when it would be needed.

During the engineering/environmental test program, an historical record of each system malfunction was maintained. Figure 7 indicates the malfunction register maintained by Reliability. The data on this register was graphed for management briefing sessions on system reliability. Effective presentations were made to management on sections within their organization which were responsible for corrective actions and their rate of progress toward resolving their particular problem areas.

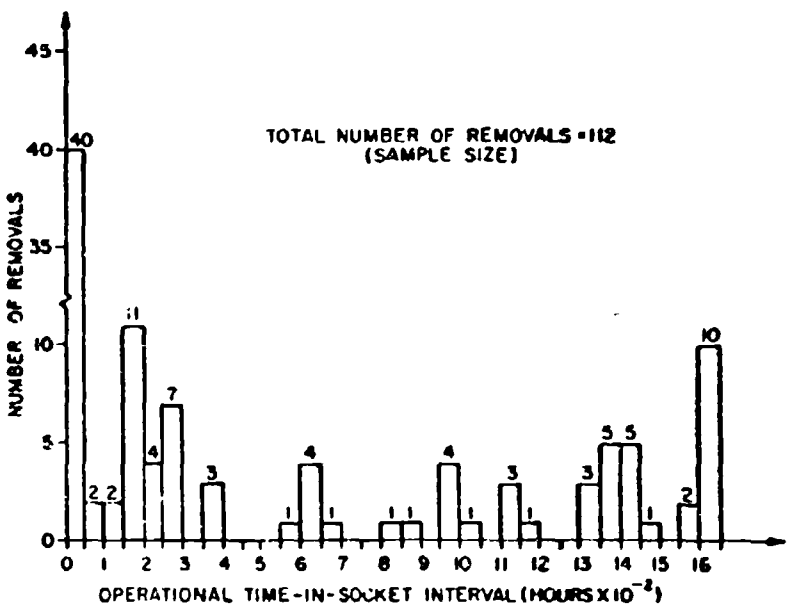


Figure 6. Number of Component Part and Electron Tube Removals/Operational Time-in-Socket Interval for Airborne Radar System No. X

MALFUNCTION REGISTER (ENGINEERING/ENVIRONMENTAL TEST PROGRAM)

ITEM NO.	*TFR NO.	SYSTEM NO.	MODULE	MOD SERIAL NO.	DESCRIPTION OF TROUBLE	CAUSE OF TROUBLE	REMARKS	SERIOUSNESS OR EFFECT	CAUSE	RESPONSIBILITY	CORRECTIVE ACTION

*TFR - TROUBLE AND FAILURE REPORT, DATA REPORTING FORM IN USE BY RELIABILITY

Figure 7. Malfunction Register (Engineering/Environmental Test Program)

THE RELIABILITY OF HERMETICALLY SEALED EQUIPMENT

By: W. B. Rosenagel
Assistant Chief Reliability Engineer
Kearfott Division of General Precision, Inc.
Little Falls, New Jersey

SUMMARY

The Kearfott Division of General Precision, Inc., has designed hundreds and manufactured thousands of hermetically sealed servomechanisms and computers. On many of these equipments, Kearfott has been the sole repair source. This enables a company to analyze all the failures without having a biased sample. A high confidence results from such a failure analysis because tampering, improperly serviced and damaged units are easily detected.

This paper reviews the failure analysis of various sealed servomechanism modules from three widely used USAF systems:

1. Central Gyro Reference System, AJA-2 (C.G.R.S.)
2. J-4 Navigational System
3. True Heading Navigation System, AJA-1

A wide range of specific MTBF's on synchros, clutches and motors, resulting from these studies, has been listed in this paper. It is believed that this may be the first time such comprehensive failure data, concerning specific operational equipment, has been published for the benefit of the industry. In addition to being the basis of a comprehensive reliability program, this data shows that failure rate correlation on the same component in the same application, but in different systems, has been as close as 1.2%.

TEXT

There are two basic prerequisites for accurate reliability data:

1. Good records of operating time
2. A system of recording and analyzing all failures occurring during this period of time

Such data provides a high confidence in the numerator and denominator and the resulting MTBF of the basic reliability equation:

$$\text{Mean-Time-Between-Failure (MTBF)} = \frac{\text{Operating Time}}{\text{Failures}}$$

Definition of MTBF

Mean-Time-Between-Failure is the total measured operating time of a population of equipments divided by the total number of failures within the population during the measured period of time. (This measurement is made during the useful life of the equipment.)

Definition of Average Life

Average life is the mean value for a normal distribution of lives of a component and generally applies to mechanical failures resulting from wearout.

It must be remembered that MTBF and average life are two separate and distinct factors. They must not be confused.

C.G.R.S. RELIABILITY ANALYSIS

The C.G.R.S. was designed by Kearfott in 1955, and manufactured at Kearfott plants for use in the USN WV-2 and USAF EC-121 Lockheed Constellation Early Warning Search Aircraft as a primary vertical reference. This system contains seven hermetically sealed servomechanism modules and the first production three-gyro, three-gimbal platform (stable element) ever manufactured by the industry. This paper covers an analysis of the failures over a four-year period on these servo cans.

We in Kearfott Reliability are fortunate to have data on each of the repaired hermetically sealed C.G.R.S. servo modules over a four-year period. A sealed unit, with the manufacturer's inspection stamp over the soldered sealing band, is the best method, under today's military procedures, of receiving for repair an untampered, undamaged module. Some of our avionic equipments, notably amplifiers and control panels, are not sealed, and, therefore, present problems when a valid analysis of failure data is attempted.

Several methods of determining flight time have been used by our reliability department. One consisted of a regression line study of 119 C.G.R.S. platform failures. In each instance the specific operating

time at failure was known, and an average value of 87 hours of flying time per aircraft per month was established. An analysis of the flight records at several East Coast operating bases was made by Kearfott field engineers. This provided an estimate 3% higher than the regression study. Lockheed later provided the total number of flying hours from their records. This figure was 7% above our computed 87 hours per month. Both these sources confirmed the regression line technique and indicated our estimate to be on the conservative side.

With this background, let us look into the failure investigation.

Figure I indicates the total failures of the seven C.G.R.S. servos during the period between January 1956 and March 1959.

Servo Module	80 P/N	Reported Failures	F*	MTBF
Integrator **	623412	210	288	5,470
Pitch and Roll**	623419	44	84	11,830
Ground Speed	623460	25	45	19,750
Voltage and Frequency	623470	16	32	17,300
Azimuth	630285	310	870	1,220

** Two modules used in each system.
* Failure rate in failures/10^6 hours

FIGURE I

Since there is insufficient space to show a detailed analysis of all the servo cans, the most complicated unit, and the unit with the lowest MTBF, the Azimuth Servo, was selected for this analysis. This unit, which contains 13 synchros, 3 motor generators and 3 magnetic clutches, has a total of more than 1250 parts, not including the wiring and sleeving. Figure II is a photograph of this unit with its cover removed.

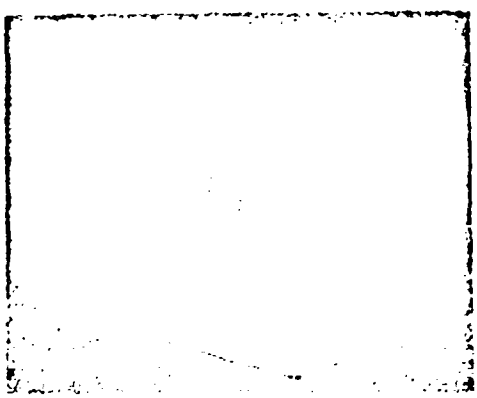


FIGURE II

Of the 310 Azimuth Servo failures reported (from Figure I), only 187 had an accurate record of operating time at failure. This time amounted to 69,424 flight hours, as reported by our field engineers under a paid program at each operating base.

The failures are broken down into two groups:

1. Unmodified Azimuth Servos - refers to initial 200 production units
2. Modified Azimuth Servos - refers to all those units produced after the first 200 and those of the unmodified units which, when rejected, were converted to modified units at time of repair. The modification constituted a change in the bearing type and the material used for several gears.

Four failure rates will be computed for the Azimuth Servo:

1. Unmodified Azimuth Servo failure rate
2. Modified Azimuth Servo failure rate
3. Actual Azimuth Servo failure rate
4. Failure rate assuming all the Azimuth Servos had been manufactured as modified units

The operating time corresponding to the failures on the modified and unmodified (Figure III) units respectively, also had to be broken into two groups. After March 1, 1957, all production Azimuth Servos became the modified type. An analysis of the C.G.R.S. delivery graph indicates how the time was apportioned. Unmodified units have been repaired at a rate indicated by curve B on this graph.

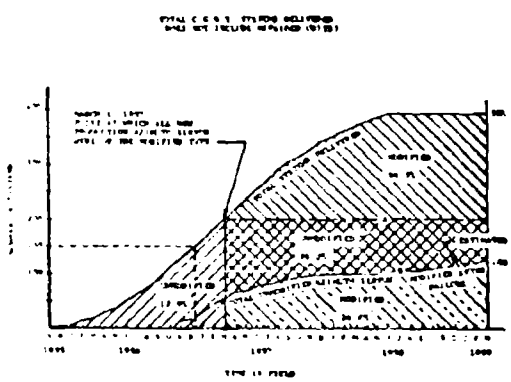


FIGURE III

By calculating the areas under the curves the cumulative operating time on unmodified and modified servos respectively is:

Unmodified Servos - 39% of total operating time

Modified Servos - 61% of total operating time

The four failure rates for the Azimuth Servo were computed to be:

Type	Failures	Operating Time, %	f	MTBF
Unmodified	143	39%	978	1020
Modified	157	61%	677	1478
Actual (combined)	310*	100%	820	1220
Theoretical	296	100%	677	1478

* Includes 16 unidentifiable units

FIGURE IV

Figure V shows the distribution of modified and unmodified can failures with time.

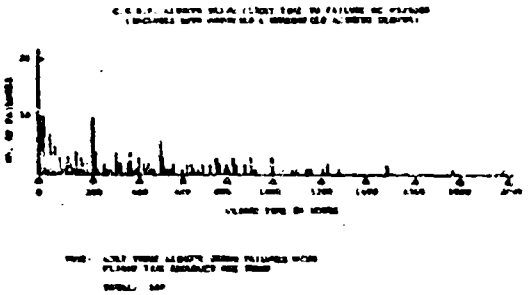


FIGURE V

Figure VI is a more detailed breakdown of the modified and unmodified can failures. Note the longest time to failure on unmodified units was 1300 hours while on the modified units it was 2000 hours indicating a growth in reliability.

Figure VII shows a component failure time comparison of the modified versus unmodified units.

Several comments must be made regarding the failure rate data in Figure VII. These failure rates are for specific circuit applications. Some of the CX's operated into single repeater loads or high impedance CT's. Others were operated into four repeaters. The gear ratios and circuit application within the Azimuth Servo

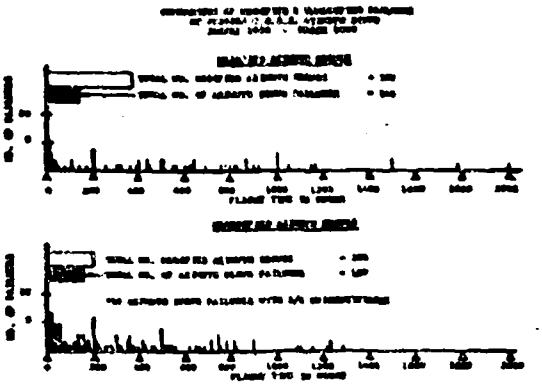


FIGURE VI

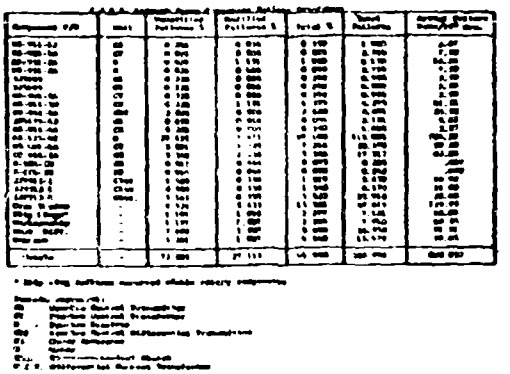


FIGURE VII

affect the mechanical stresses, notably on the ball bearings.

Since conservative estimates were made whenever necessary, we have a high confidence that the over-all MTBF was not less than that indicated. It may be considerably higher.

An arbitrary confidence of 90% has been placed on this data by our department.

J-4 RELIABILITY ANALYSIS

The J-4 Compass System incorporates an Azimuth Servo as part of the Servo Amplifier. This system, which was designed and manufactured by Kearfott, incorporates two motors and four synchros within the Azimuth Servo.

A total of 4200 J-4 Azimuth Servos have been produced by Kearfott. Early in 1959, Kearfott field engineers visited the following USAF bases to obtain failure data on the J-4: Tyndall, Suffolk McGuire,

Andrews, Seymour Johnson and Youngstown Air Force Bases. This survey indicated a total J-4 operating time on F-100, F-101, F-102 and F-104 aircraft of 83,664 hours with 109 Servo Amplifier failures. A J-4 Servo Amplifier MTBF based on this data was computed to be 755 hours or a failure rate of 1310 failures/10^6 hours. The Azimuth Servo MTBF was 845 hours or a failure rate of 1184 failures/10^6 hours.

Because 98% of the repair work to this servo had been done by other companies, complete information was not available to Kearfott. However, accurate data regarding the field repairs made by Kearfott were recorded. It was possible to extrapolate rotary component failure rates within the Azimuth Servo, using the Servo Amplifier MTBF, with reasonable confidence. The component failure rates shown in Figure VIII were derived from this study.

Component	P/N	Application	%	MTBF Hrs.
Slew Motor	325395-1	Slewing	0.1	1,175
Motor Generator	S-809-20	Mag. Hdg.	9.0	11,180
Control Transmitter	85-911-1A	Repeater Data	10.4	10,370
Control Transmitter	204173-13	Precise Data	10.4	10,370
Control Transformer	85-901-1A	Mag. Hdg.	10.4	10,370
Diff. Cont. Xfmr.	424387-1	Gyro CX Loop*	1.0	83,000
"Total"	-	-	100.0	

* One failure assumed for statistical purposes

FIGURE VIII

It is significant to note that there were no mechanical failures of gear trains, ball bearings, etc. in this analysis.

Figure IX graphically indicates the J-4 Azimuth Servo component failure rates.

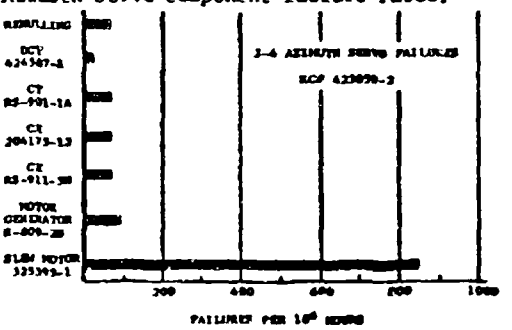


FIGURE IX

The J-4 Azimuth Servo has a high temperature rise due to its location inside the enclosed J-4 Amplifier. Actual thermocouple recorded temperature readings indicate a 43-degree C temperature rise between external ambient and the J-4 Azimuth Servo can temperature. This fact of high temperature operation must be considered in analysing the failure rates.

True Heading (AJA-1) Reliability Analysis

The True Heading System was designed and manufactured by Kearfott for the Boeing B-52 Model E, F and G "Stratofortress".

The system, used in conjunction with the N-1 Navigational System, consists of three modules:

1. Master Indicator Class II
2. True Heading Indicator
3. True Heading Computer Amplifier

Items 1 and 2 and the computer part of item 3 are hermetically sealed.

The actual system and module reliability figures are classified and consequently the analysis detail is omitted, but a partial list of the component failure rates can be released.

The period covered by the analysis extended from January 1, 1958 through July 1, 1959. The data inputs were provided by three separate sources. Boeing Airplane Company supplied the flight time. IBM supplied the total number of failures. Kearfott analyzed the failures after repair of the hermetically sealed units. Since the data came from three different reliability organizations, the confidence level of the resulting component failure rates is estimated to be 10% less than that for the C.G.R.S.

In Figure X, the failure rates represent a partial list of the AJA-1 component types and applications similar to those previously detailed on the other two systems.

Component P/N	Unit	Application	%	MTBF
85-911-1A	CX	V. Hdg. and Mag. Var.	21.6	46,296
85-911-6001	CX	Lat. Corr.	23.3	43,103
85-911-4A	CX	Lat. Corr.	12.5	80,000
S-710-2A	CX	Repeater Data	85.2	11,737
6152740	CX	Precise Data	12.3	80,306
85-901-1E	CDX	True Heading	23.2	43,103
85-901-1A	CT	Longitude	23.2	43,103
424387-1A	CT	True Heading	88.3	11,327
424387-1A	CT	Mag. Var.	172.0	5,787
19951-6001	CT	Azimuth	170.2	5,873
S-809-20	MG	True Heading	64.7	15,456
S-809-20	MG	Mag. Var.	349.8	2,892
S-119-20	M	Azimuth	316.0	3,165

FIGURE X

Differences in severity of application are most marked for the CT's and motor generators.

CONCLUSIONS

The RS-901-1A Control Transformer (CT) Synchro used in the magnetic heading (Mag. Hdg.) loop of the C.G.R.S. and the identical unit in the same loop of the

J-4 show remarkable correlation:

RS-901-1A	f	MTBF
CGRS Mag. Hdg. Loop	59.20	16,892
J-4 Mag. Hdg. Loop	60.4	16,556

These failure rates for the same application of the Kearfott RS-901-1A CT, but in two different systems, correlate within 1.2%.

Circuit application and severity of operation strongly influence the reliability of an electromechanical component. For example, two motor generators of the same type (R809-2B), used in two different servo loops of the AJA-1 system, have widely differing MTBF's. In the True Heading loop the motor generator has an MTBF of 15,456 hours, while the same component in the Magnetic Variation loop has an MTBF of 2,892. The reason for this significant difference is that the Magnetic Variation loop has a higher ambient temperature (because of location), a higher generator excitation voltage (115 volts as compared to 26 volts), a higher gear ratio and the generator excitation is keyed to reduce damping and provide a "fast slew" during initial system set-up.

Figure XI is a summary of synchro MTBF's used in the three systems discussed in this article. The units are listed in accordance with the order of increasing MTBF's. It should be noted that some of control transmitters (CX) with the highest MTBF's were used in circuits where no loads were connected to the stators.

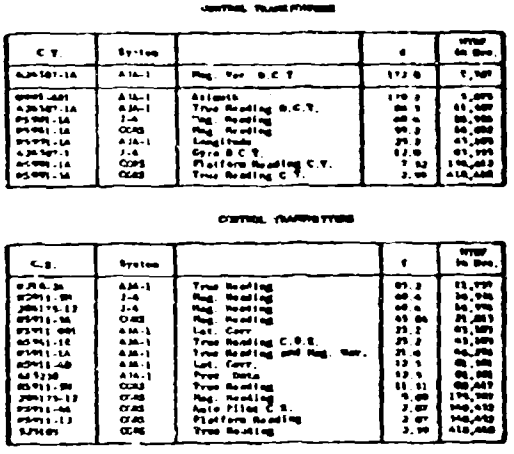


FIGURE XI

With good reliability field data as a basis, in-plant life tests under various degrees of thermal, electrical and mechanical stress levels have been initiated on all basic motor, motor generator and synchro types. The results of these tests have been used, along with field data, to determine "K" factors. A "K" factor is a reliability correction coefficient for various environments, such as laboratory conditions, shipboard, trailer mounted, manned aircraft and missile applications. In the case of manned aircraft, we have been able to determine specific "K" factors for particular aircraft types and branch of services.

Accurate secondary data, such as described in this paper, has enabled Kearfott to predict reliability in new systems with startling confidence.

Careful review of reliability data and its expeditious feedback to the designers has enabled the component and systems engineers to take prompt remedial action to improve all new system designs. This data is also used as the basis for spare parts provisioning programs, cost reduction and product improvement programs.

Remember, the best reliability data in the world is no good in a file; it must be transmitted to, and applied by, the designers of new equipment.

A MEASURE OF RELIABILITY AND INFORMATION QUALITY IN REDUNDANT SYSTEMS

By: Stanley A. Rosenthal, Herbert Jaffe, and Mordecai D. Katz
Air Armament Division, Sperry Gyroscope Company
Division of Sperry Rand Corporation
Great Neck, New York

In the development of complex military electronic equipment it is necessary to consider the effects of many interacting operational and design factors. For both the overall weapon system, and for the electronic subsystems with which we are concerned, optimum trade-offs for such interdependent parameters as reliability, weight, accuracy, maintainability, cost, mission time, ar target vulnerability must be determined. Considered individually, these parameters do not adequately describe the operational worth of equipment incorporating multiple redundancies and operating modes. Operational worth is the probability that a system will achieve success in performing its required functions, and is obviously a prime consideration in the planning and selection of systems for military application.

The inadequacy in describing operational worth in terms of its individual parameters was encountered in the synthesis of an optimum redundant configuration for the B-58 bombing-navigation system. To resolve this difficulty, operational worth models describing the interaction between parameters were developed and utilized by Sperry Gyroscope Company. These techniques enabled Sperry to design this bombing-navigation system to have a high operational worth and an inherent mission reliability in excess of 95 per cent. In this paper, the model for "system worth", the inter-action between reliability and information quality, is discussed.

THE B-58 BOMBING-NAVIGATION SYSTEM

The B-58 bombing-navigation system is a Doppler-stellar-inertial system which provides highly accurate guidance of the supersonic aircraft to its target, and automatic bombing. Continuous determination of such parameters as aircraft position and heading, groundspeed, steering, ground track, distance to destination, attitude, altitude, true airspeed, wind velocity, and bombing data is accomplished through the use of a complex combination of sensors and electronic-mechanical analog computers. The computer receives input data from the sensors which include search and Doppler radars for position and velocity monitoring, inertial reference units for velocity and attitude determination, a celestial tracker for stellar navigation, as well as a radio altimeter and a magnetic heading reference. The computer equipment performs such diverse functions as determination of aircraft position, great-circle steering, computation of bomb ballistics data, latitude and longitude integration, and solution of the radar sighting problem. Figure 1 shows the installation in the aircraft.

Figure 2 is a pictorial representation of the complete B-58 bombing-navigation system showing data flow between major components of the system.

By their nature, complex systems have a low probability of completing a mission without any malfunctions. To attempt reliability improvement merely by reducing complexity would sacrifice much of the potential utility of the equipment.

High reliability in the B-58 system is achieved by means of two types of malfunction protection. Redundant or multipurpose circuits and components are used wherever practical, consistent with the requirement of minimum weight for airborne equipment. In addition, alternate methods of operation are provided so that successful completion of a mission may be accomplished in any one of several different system operating modes. Although these techniques yield substantial improvements in reliability with relatively little additional equipment, their use generally involves the replacement of primary circuit equipment with sensing and computing elements of lesser accuracy.

Examples of the types of redundancies and alternate operational capabilities which have been incorporated in this system are:

Redundant replacements for -

- Primary inertial stabilization system,
- Stellar computer electronics,
- Heading circuit electronics.

Multi-purpose redundant replacements for -

- Relative heading, course angle, ground track, true heading, aim-point bearing, and sighting angle servos (a single servo malfunction assembly can be switched to replace any of the six servos),
- Time-to-go, altitude above sea level, and sighting computers (a single auxiliary computer can be switched to replace any one of the three computers).

Alternate capabilities for -

- Search radar position correction (Astro tracker position fix),
- Astro-tracker heading determination (Inertial or magnetic heading),

Doppler radar velocity reference
(Inertial velocity or airspeed reference).

Complementing these major reliability provisions is a complete in-flight malfunction tester for automatic checkout of the system to determine whether use of replacement equipment is indicated. In addition, a malfunction control panel enables the navigator to switch to alternate system modes so that the aircraft can complete its mission.

In the following sections the concept and defining equations for system worth are developed and an illustrative computation is presented. The accuracies and reliabilities for this illustration are for a hypothetical configuration and do not reflect the actual configuration or capability of the B-58 bombing-navigation system.

THE CONCEPT OF SYSTEM WORTH

In order to optimize the trade-off between reliability and information quality for a fixed design constraint, such as total system cost or weight, an interaction model for system worth was developed. System worth is defined as the probability that the computed output of a system with redundant modes will be within a specified quality considering all possible satisfactory modes of operation. In the B-58 system, the measure of information quality is taken to be the accuracy with which navigation and bombing functions are performed.

In developing a mathematical model for these satisfactory modes of operation, the system was divided into functional blocks (Figure 3). Each block is a portion of the system which performs a specific function but does not include any redundancy or other malfunction replacement. Malfunction mode initiation, therefore, consists of switching entire functional blocks or groups of functional blocks into various operating configurations.

The probability of all primary equipment operating satisfactorily from take off to bomb release is defined as the normal mode probability, and is expressed as:

$$P_{R_o} = R_1 \cdot R_2 \cdot R_3 \cdots R_n = \prod_{i=1}^{n} R_i \qquad (1)$$

where $R_1, R_2, - - - - - R_n$ are the reliabilities of the individual functional blocks constituting the normal mode.

In the event of a malfunction affecting normal mode system operation, the applicable replacement circuitry is employed and the mission is completed in an off-normal mode configuration. The off-normal modes of operation are mutually exclusive; their probabilities of occurrence are computed as the product of the reliabilities of the functional blocks constituting the particular off-normal mode and the probability of the occurrence of the failure or failures defining the mode:

$$P_{R_k} = (1-R_k)R'_k \cdot \prod_{\substack{i=1 \\ i \neq k}}^{n} R_i = (1-R_k)R'_k\left(\frac{P_{R_o}}{R_k}\right) \qquad (2)$$

where $(1 - R_k)$ = probability of the k^{th} functional block failing during the mission,

and R'_k = reliability of the replacement circuitry of the k^{th} block.

Similarly, the probability of occurrence of a multiple malfunction mode may be determined from:

$$P_{R_{s,t,\cdots z}} = (1-R_s)(1-R_t)\cdots(1-R_z)R'_sR'_t\cdots R'_z \cdot \prod_{\substack{i=1 \\ i \neq s,t,\cdots z}}^{n} R_i \qquad (3)$$

Figures 4 and 5 illustrate these relationships.

Computational accuracy for the system is expressed by a normal bivariate distribution and is specified in terms of a circular probable error (CPE) for each system mode. Associated with each of these system modes are computational errors for such system functions as dead reckoning, radar sighting, and bombing. These mode errors are obtained by an appropriate root-mean-square (RMS) statistical combination of the inaccuracies of the individual components utilized in that mode configuration.

Optimum accuracy for the system is achieved in the normal mode of equipment operation, that is, with all equipment operating properly, as well as in those modes utilizing redundant replacement circuitry of equivalent accuracy. Reduced accuracy, resulting in larger system errors, is experienced in those modes where alternate sensing and computing methods of lesser accuracy are employed.

Consequently, successful completion of a mission can be accomplished either with normal accuracy, in the normal or redundant replacement modes, or with degraded accuracy in one of the alternate capability modes.

For the assumption of a normal bivariate distribution of system errors centered at the target with equal and independent range and cross-range standard deviations, the probability distribution of the system experiencing a radial error ρ is:

$$p(\rho) = \frac{\rho}{\sigma^2} e^{-\frac{\rho^2}{2\sigma^2}} \qquad (4)$$

where ρ = system radial error

$\sigma = \sigma_x = \sigma_y$

The probability that an error ρ lies within a circle of radius r is then obtained by integrating the function of Equation 4:

$$P_{A_j}(\rho \le r) = \int_0^r \frac{\rho}{\sigma_j^2} e^{\frac{-\rho^2}{2\sigma_j^2}} d\rho = 1 - e^{\frac{-r^2}{2\sigma_j^2}} \quad (5)$$

where P_{A_j} = jth mode accuracy probability (cumulative error distribution for the jth mode)

r = specification radial error value

σ_j = jth mode RMS error

Circular probable error, for a bomb strike, is defined as the radius of a circle, centered about the desired target point, which has a 50% probability of encompassing the actual burst point. If $P_{A_j}(\rho \le r)$ is taken as 0.5 in Equation 5, then the magnitude of CPE radius in terms of system rms error in range or cross range σ_j is:

$$r_{CPE_j} = 1.177\,\sigma_j \quad (6)$$

The system rms radial error σ_{A_j} is the square root of the sum of the squares of the rms errors in range and cross range attributable to each error source:

$$\sigma_{A_j} = \sqrt{\sigma_x^2 + \sigma_y^2} = \sqrt{2}\,\sigma_j \quad (7)$$

and

$$r_{CPE_j} = \frac{1.177\,\sigma_{A_j}}{\sqrt{2}} = 0.832\,\sigma_{A_j} \quad (8)$$

Consequently the accuracy probability is:

$$P_{A_j} = 1 - e^{\frac{-r^2}{\sigma_{A_j}^2}} \quad (9)$$

The product of the mode cumulative-error-distribution function and its respective mode occurrence probability yields a continuous mode system worth increment:

$$P_j(\rho \le r) = P_{R_j} P_{A_j} \quad (10)$$

or

$$P_j(\rho \le r) = P_{R_j}\left[1 - e^{-\left(\frac{r}{\sigma_{A_j}}\right)^2}\right] \quad (11)$$

where P_{R_j} = j^{th} mode occurrence probability (probability of operating successfully in the j^{th} mode)

The sum of these increments for all satisfactory modes of system operation gives the total probability of delivering a bomb to within the specified distance r of the target. This total probability is the system worth:

$$P(\rho \le r) = \sum_{j=1}^{n} P_{R_j}\left[1 - e^{-\left(\frac{r}{\sigma_{A_j}}\right)^2}\right] \quad (12)$$

In terms of CPE, from Equation 8:

$$P(\rho \le r) = \sum_{j=1}^{n} P_{R_j}\left[1 - e^{-\left(\frac{0.832r}{CPE_j}\right)^2}\right] \quad (13)$$

Plotting Equation 13 yields a monotonically increasing system worth curve of the type shown in Figure 6. The shape of the curve is a function of the accuracies of the individual system modes and their respective probabilities of occurrence. The asymptote is the sum of the mode occurrence probabilities for all satisfactory operating modes, that is, the probability of completing a mission in one of the acceptable system configurations. This probability is the familiar definition of system reliability.

Overall system worth is affected by changing either reliability or accuracy. A trade-off between them is possible and, because of their conflicting nature, generally necessary. During system synthesis, it is desirable to develop system worth curves for a graphical comparison of the effects of different system design criteria. Satisfaction of a given CPE system specification requirement can be achieved with various combinations of mode reliabilities and their respective accuracies, as illustrated in Figure 7.

In analyzing the performance of the B-58 bombing-navigation system, probability equations for over five hundred of the most probable operating modes have been programmed for UNIVAC® computation. Thus, the effects on system worth due to changes in functional block reliability or error distribution are rapidly and accurately obtained from the computer for engineering evaluation.

EXAMPLE OF SYSTEM WORTH APPLICATION

As an illustration of the application of system worth, consider the following example. Figure 8 is a functional block diagram of a hypothetical, greatly simplified bombing system. This system incorporates a redundant bombing computer, a heading sensor, and multiple sensing capabilities for measuring position and groundspeed. The normal mode of system operation utilizes a tracking radar for ground position fixing, a star tracker for heading reference, a Doppler radar for groundspeed determination, and a primary bombing computer for generating a bomb-release signal. Failure of any normal mode functional block necessitates switching to a different off-normal system configuration. In this example, there are seven possible off-normal modes of successful system operation, as follows:

Single Malfunctions

- Tracking Radar - Position is determined by using the star tracker for a celestial fix.
- Doppler Groundspeed Radar - Airspeed plus wind memory is used for groundspeed determination.
- Computer - An identical redundant bombing computer is available.

Double Malfunctions

- Tracking Radar and Doppler Groundspeed Radar.
- Tracking Radar and Computer.
- Doppler Groundspeed Radar and Computer.

Triple Malfunction

- Tracking Radar, Doppler Groundspeed Radar, and Computer.

The redundant bombing computer is assumed to be as accurate as the primary computer; consequently, there is no degradation in system accuracy in the computer malfunction mode. The reliability of the bombing computer subsystem can therefore be taken as the composite probability of either the primary computer or the redundant computer available for use:

$$R_4^* = 1 - (1-R_4)(1-R'_4) \qquad (14)$$

This effectively reduces the number of off-normal modes from seven to three as shown in Table I. Associated with the normal mode and with each of the three off-normal modes listed in this tabulation is its probability of occurrence, the bombing accuracy of the system when operating in each configuration, and the incremental as well as total system worth for various error radii. Figure 9 is the resultant system worth curve for this example.

Consider now that a specified weight increase is allowed for a system worth improvement. This weight allowance might be applied to the system in several ways. Three possibilities and their effects on system worth will be evaluated:

1. Improvement of the reliability of a functional block; e.g., through internal redundancy, the tracking radar reliability is improved from .75 to .96.
2. Incorporation of additional alternate mode capabilities; e.g., addition of a flux valve as a secondary heading reference and a navigation satellite receptor as a tertiary position fix backup. The resulting alternate modes are less accurate than the normal mode (Figure 10).
3. Improvement in mode accuracies; e.g., larger, more precise components are used to reduce errors in sensing and computing.

System worth results for each of these configurations are shown in Tables II, III, and IV respectively and are plotted as curves II, III, and IV in Figure 11. For comparison, the system worth curve of the original configuration is also included in this graph as curve I.

The improvement of the reliability of the tracking radar results in an increase in the probability of being in the normal mode of operation. The system worth is improved over the entire range of miss-distances, with the largest improvement being in the 50-unit to 300-unit miss-distance range.

The incorporation of additional alternate mode capabilities of lesser accuracies than the primary modes results in improved system reliability with little improvement in system worth below 100 units of miss-distance. For greater miss-distances, the probability of a successful bomb delivery (the system worth) is increased due to the higher composite reliability of the heading reference and position-fixing subsystems.

The improvement in mode accuracies yields the greatest system worth benefits for small miss-distances with no increase whatsoever in system reliability.

The decision of which of these improvements should be incorporated in the system design must

then be evaluated on the basis of such other parameters as cost, maintainability, producibility, target vulnerability, mission time, and the radius of destruction of the warhead.

CONCLUSION

The system worth concept is an essential tool in the synthesis and analysis of complex systems. As described, it enables the system designer to evaluate the incremental benefits attributable to specific system improvements in accuracy, reliability, redundancy, and alternate mode capability. During the early phases of system design, the accuracy and reliability data used in the system worth analysis will be predicted from empirical relationships. Verification of the empirical data can later be made from field data when the system becomes operational.

Although the system worth model was developed for the synthesis of an optimum redundant configuration for the B-58 bombing-navigation system, it has universal application to any system where alternate paths of differing information-handling capability are used.

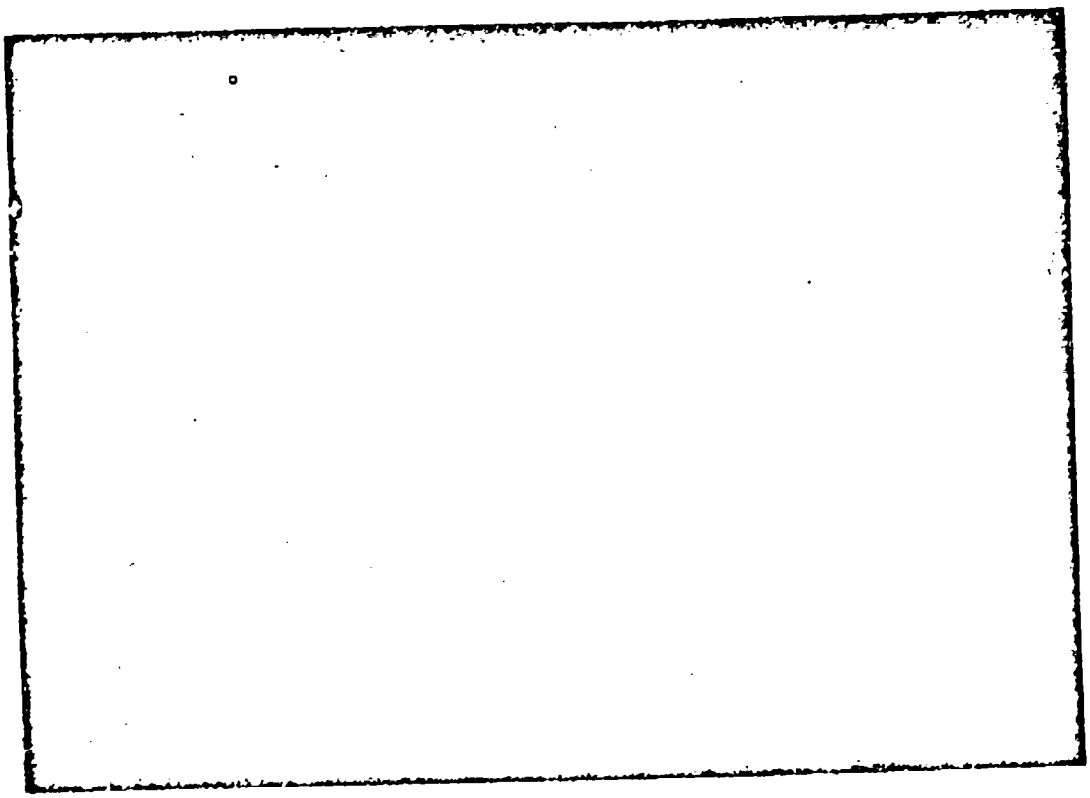


TABLE I

SYSTEM WORTH ANALYSIS

FOR ORIGINAL PROTECTED CONFIGURATION

	OPERATING MODE	ACCURACY (UNITS OF MISS-DISTANCE) r_{CPE_j}	PROBABILITY OF BEING IN THE MODE P_{R_j}	SYSTEM WORTH FOR VARIOUS RADII OF ERROR CIRCLE $P = P_{R_j}\left[1-e^{-(0.832r/r_{CPE_j})^2}\right]$						
				r=10	25	50	100	250	500	1500
	NORMAL	50	$R_1R_2R_3R_4^*$ = .536	.014	.085	.267	.502	.536	.536	.536
OFF-NORMAL	TRACKING RADAR MALFUNCTION	200	$(1-R_1)R_2R_3R_4^*$ = .179	.000	.002	.008	.028	.113	.176	.179
OFF-NORMAL	DOPPLER GROUNDSPEED RADAR MALFUNCTION	75	$(1-R_2)R_2'R_1R_2R_4^*$ = .090	.001	.006	.024	.063	.089	.090	.090
OFF-NORMAL	TRACKING & DOPPLER RADARS MALFUNCTIONS	225	$(1-R_1)(1-R_3)R_2R_3'R_4^*$ = .030	.000	.000	.001	.004	.017	.029	.030
	TOTAL SYSTEM WORTH			.015	.093	.300	.597	.755	.831	.835

TABLE II

SYSTEM WORTH ANALYSIS

FOR IMPROVED TRACKING RADAR RELIABILITY

	OPERATING MODE	ACCURACY (UNITS OF MISS-DISTANCE) r_{CPE_j}	PROBABILITY OF BEING IN THE MODE P_{R_j}	SYSTEM WORTH FOR VARIOUS RADII OF ERROR CIRCLE $P = P_{R_j}\left[1-e^{-(0.832r/r_{CPE_j})^2}\right]$						
				r=10	25	50	100	250	500	1500
	NORMAL	50	$R_1R_2R_3R_4^*$ = .679	.018	.107	.339	.636	.679	.679	.679
OFF-NORMAL	TRACKING RADAR MALFUNCTION	200	$(1-R_1)R_2R_3R_4^*$ = .036	.000	.000	.002	.005	.024	.035	.036
OFF-NORMAL	DOPPLER GROUND SPEED RADAR MALFUNCTION	75	$(1-R_2)R_3'R_1R_2R_4^*$ = .114	.001	.006	.030	.060	.113	.114	.114
OFF-NORMAL	TRACKING & DOPPLER RADARS MALFUNCTIONS	225	$(1-R_1)(1-R_3)R_2R_3'R_4^*$ = .006	.000	.000	.000	.000	.003	.005	.006
	TOTAL SYSTEM WORTH			.019	.115	.371	.721	.819	.835	.835

TABLE III

SYSTEM WORTH ANALYSIS

ADDITION OF FLUX VALVE AND NAVIGATION SATELLITE RECEPTOR

	OPERATING MODE	ACCURACY (UNITS OF MEAN-DISTANCE, r_{CPE_j})	PROBABILITY OF BEING IN THE MODE P_{R_j}	SYSTEM WORTH FOR VARIOUS RADII OF ERROR CIRCLE $P = P_{R_j}\left[1-e^{-(0.832r/r_{CPE_j})^2}\right]$						
				r=10	25	50	100	250	500	1500
	NORMAL	50	$R_1R_2R_3R_4^*$ = .536	.014	.085	.267	.502	.536	.536	.536
OFF-NORMAL	TRACKING RADAR MALFUNCTION	200	$(1-R_1)R_2R_3R_4^*$ = .179	.000	.002	.008	.028	.119	.176	.179
OFF-NORMAL	STAR TRACKER MALFUNCTION	150	$(1-R_2)R_2'R_1R_3R_4^*$ = .094	.000	.002	.007	.024	.081	.094	.094
OFF-NORMAL	DOPPLER RADAR MALFUNCTION	75	$(1-R_3)R_3'R_1R_2R_4^*$ = .090	.001	.006	.024	.063	.089	.090	.090
OFF-NORMAL	TRACKING RADAR & STAR TRACKER MALFUNCTIONS	250	$(1-R_1)(1-R_2)R_1^*R_2'R_3R_4^*$ .027	.000	.000	.000	.002	.013	.025	.026
OFF-NORMAL	TRACKING RADAR & DOPPLER RADAR MALFUNCTIONS	225	$(1-R_1)(1-R_3)R_3'R_2R_4^*$ = .030	.000	.000	.001	.004	.017	.029	.030
OFF-NORMAL	STAR TRACKER & DOPPLER RADAR MALFUNCTIONS	175	$(1-R_2)(1-R_3)R_2'R_3'R_1R_4^*$ .016	.000	.000	.001	.003	.012	.016	.016
OFF-NORMAL	TRACKING RADAR & STAR TRACKER & DOPPLER RADAR MALFUNCTIONS	273	$(1-R_1)(1-R_2)(1-R_3)R_1'R_2'R_3'R_4^*$ = .004	.000	.000	.000	.000	.001	.003	.003
	TOTAL SYSTEM WORTH			.015	.095	.308	.626	.867	.966	.971

TABLE IV

SYSTEM WORTH ANALYSIS
FOR IMPROVED MODE ACCURACY

OPERATING MODE		ACCURACY (UNITS OF MISS-DISTANCE) r_{CPE_j}	PROBABILITY OF BEING IN THE MODE P_{R_j}	SYSTEM WORTH FOR VARIOUS RADII OF ERROR CIRCLE $P = P_{R_j}\left[1-e^{-(0.832r/r_{CPE_j})^2}\right]$						
				r=10	25	50	100	250	500	1500
NORMAL		35	$R_1R_2R_3R_4^c$ .536	.030	.158	.406	.532	.536	.536	.536
OFF-NORMAL	TRACKING RADAR MALFUNCTION	200	$(1-R_1)R_2R_3R_4^c$ = .179	.000	.002	.008	.028	.119	.176	.179
	DOPPLER GROUND SPEED RADAR MALFUNCTION	65	$(1-R_3)R_3'R_1R_2R_4^c$ = .090	.001	.008	.030	.072	.090	.090	.090
	TRACKING & DOPPLER RADAR MALFUNCTION	225	$(1-R_1)(1-R_3)R_2R_3'R_4^c$ = .030	.000	.000	.001	.004	.017	.029	.030
TOTAL SYSTEM WORTH				.031	.168	.444	.636	.762	.831	.835

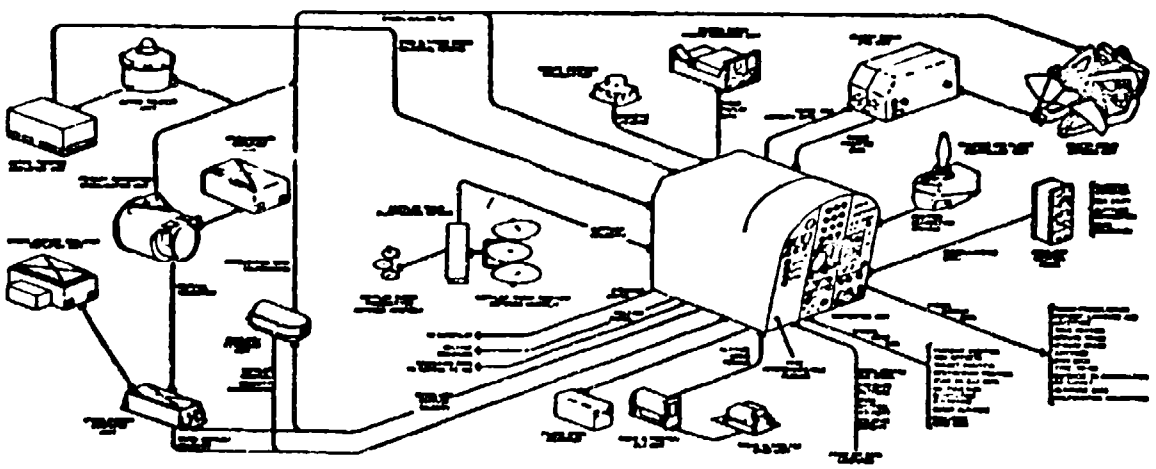


FIG 2-PICTORIAL REPRESENTATION OF [illegible]

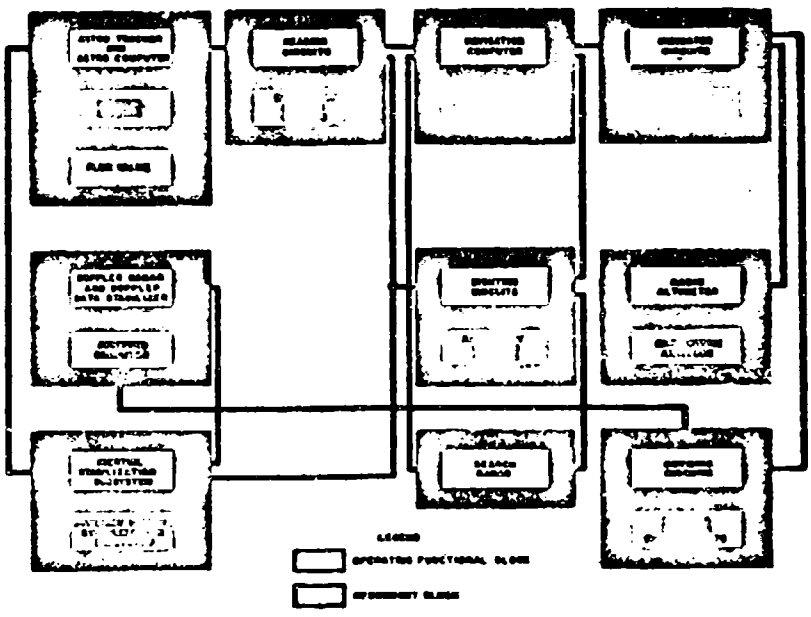


FIG. 3
SIMPLIFIED FUNCTIONAL DIAGRAM OF B-58 BOMBING-NAVIGATION SYSTEM

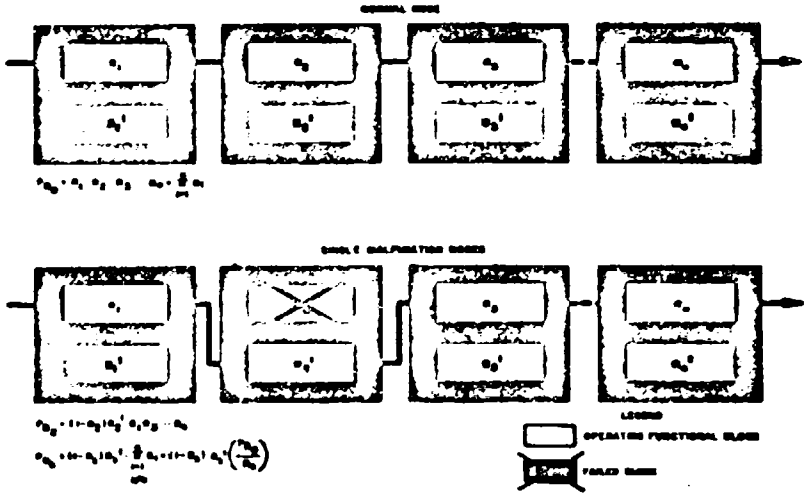


FIG. 4
NORMAL MODE AND SINGLE MALFUNCTION MODES

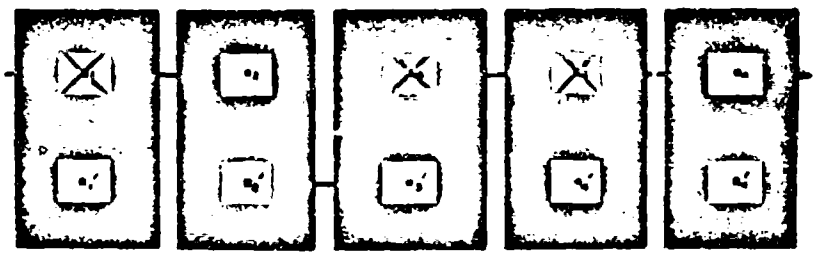


FIG. 5
MULTIPLE MALFUNCTION MODES

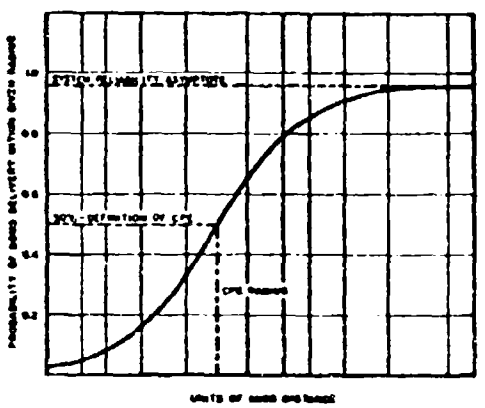


FIG. 6
SYSTEM WORTH AS A MEASURE OF RELIABILITY AND ACCURACY

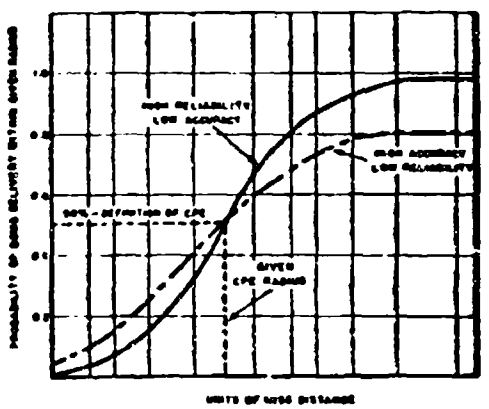


FIG. 7
SATISFACTION OF GIVEN CPE REQUIREMENTS WITH DIFFERENT SYSTEM WORTH FUNCTIONS

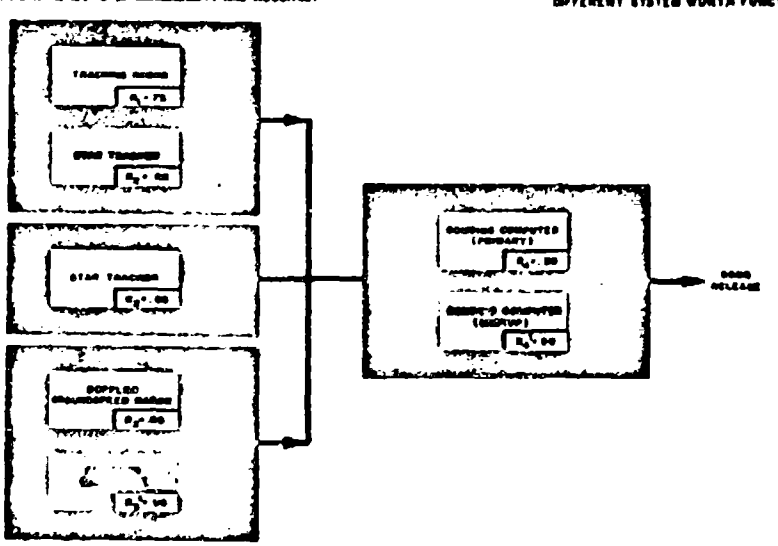


FIG. 8
FUNCTIONAL DIAGRAM OF SIMPLIFIED BOMBING SYSTEM

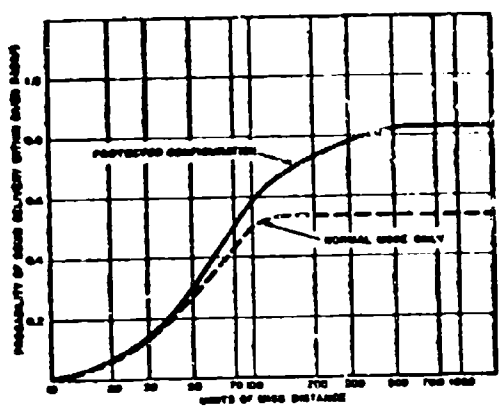


FIG. 9
SYSTEM WORTH OF HYPOTHETICAL BOMBING SYSTEM
(SHOWING COMPARISON WITH UNPROTECTED NORMAL MODE)

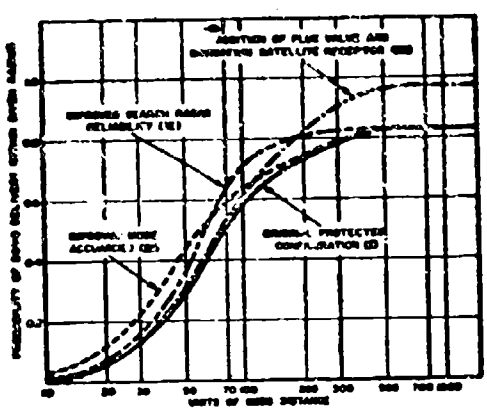


FIG. 10
COMPARISON OF SYSTEM WORTH IMPROVEMENTS
FOR HYPOTHETICAL BOMBING SYSTEM

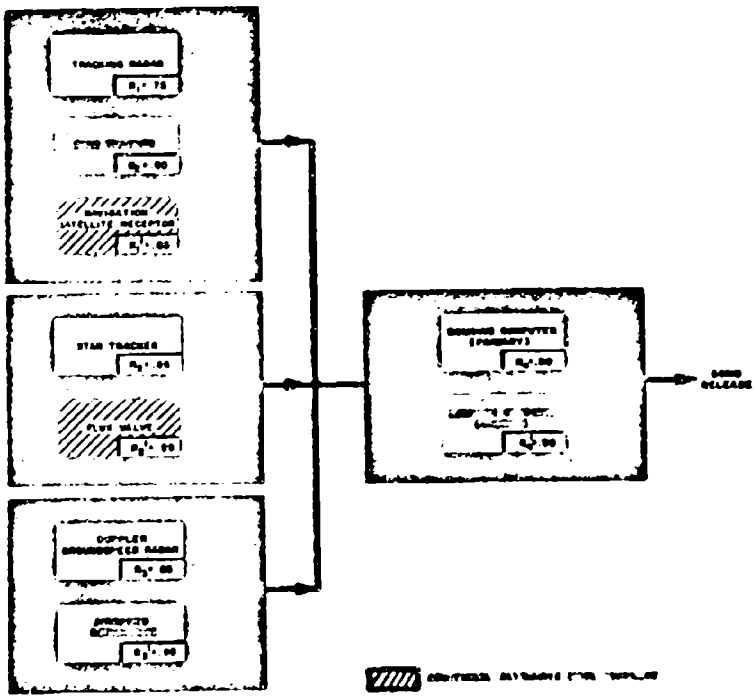


FIG. 11
FUNCTIONAL DIAGRAM OF SIMPLIFIED BOMBING [illegible]

ELECTROMAGNETIC ENVIRONMENT SIMULATION FOR SYSTEM TRAINERS

By: F. P. Cullen, W. Rolf Jr., Stromberg-Carlson, A Division of General Dynamics Corporation and J. K. Scully, Hertronics

INTRODUCTION

The complexity of modern airborne electromagnetic surveillance and countermeasures systems requires an intensive training program in order to achieve a high order of operator proficiency. Economic and other considerations preclude the operator from attaining this proficiency level by actual mission experience, therefore, use must be made of simulator devices that can present with a high degree of realism the conditions that the operator would experience during an actual mission.

The usual approach to this simulation problem has been by the use of digitally coded punched paper or magnetic tape program which provides command information for the introduction and removal of the simulated electromagnetic environment stimuli. The use of this technique has the disadvantage of (1) requiring extensive facilities for the preparing of the tape programs, (2) difficult to change mission goal during a programmed run since it usually means a complete shut down and the substitution and preparation of a new tape program, (3) usually not possible to run an integrated training mission with other systems that may be aboard the vehicle in question, (4) this approach inevitably requires the development of a completely different simulator for each operational system in question resulting in more development dollars being expended by the training commands.

This paper shall present a scheme for providing a simulator that is universal in nature without having imposed upon it the restrictions that are noted above. It provides a means for providing the electromagnetic environment from programmed navigational data, it is adaptable to integrated system operation and it can be utilized with a greater number of different reconnaissance and electronic countermeasures systems by simple expediency of providing modular adapters to adapt to the various configurations thereby reducing the development cost considerable since the basic simulator is based upon the actual geographic location of the electromagnetic environment.

SIMULATION PROBLEM

The problem is basically one of providing indications of a detected environment which will approach in realism the order and sequence of detection that occurs during an actual mission. The magnitude and distribution of these indications must be governed by all factors which influence the air vehicles detection system. These factors include vehicle velocity, heading, altitude, emitter power and emitter antenna coverage. The principle upon which this simulator is based is to divide the area of air vehicle operation into a number geographic quadrants. Each quadrant will have assigned to it the capability of generating a number and variety of indications corresponding to the types of RF emitters contained in that quadrant. The method of generation of this simulated environment and the design considerations necessary to achieve realism are discussed in the following sections.

Since reliability has been described as an inverse function of complexity, the devices and circuits utilized to implement this technique have been selected to achieve a high degree of realism while maintaining a relatively low order of complexity.

DESCRIPTION OF THE SIMULATOR

A functional block diagram of the system is given in figure 1. Positional data inputs defining latitude and longitude may be introduced in either analog or digital format since a digital to analog converter is provided in addition to the position synchros. There are two position synchros, one for latitude and one for longitude information. These synchros will control the selector commutator servos which in turn activate the quadrant relays.

The output of the synchro unit provides data to the two operational groups of the simulator. These groups are the directional coding section and the environment activating group.

Environment activation is accomplished by the quadrant selection unit which determines the areas of the environment selection unit to be energized. The quadrant selection unit consists of a relay matrix with the rows defining increments of latitude and the columns representing increments of longitude. Connections to a quadrant relay are made only at those intersections where a known electromagnetic environment exists. The commutator segments are arranged to energize five latitude and five longitude lines of the matrix. The particular group of relays which is activated programs the output of the environment selection unit.

The function of the other section, the directional coding section, is to process the output of the synchro unit and generate a binary description of direction relative to the air vehicle. The direction and rate of shaft rotation of the synchros is sensed by the differential heading encoder. This unit makes a comparison of the direction and rate of change of the analog output of the position synchros. A binary description of heading is generated which is used in the directional logic unit to interpret the individual quadrant direction

codes relative to the air vehicle heading. The output from the directional logic consists of a four bit binary word which defines the simulated emitter's position relative to the air vehicle heading.

The environment selection unit (Fig. 2) contains a high power environment sector activity switch associated with each of a possible 25 quadrant relays which may be activated in the relay matrix. The quadrant relays determine which level of the switch is to be energized. At any level, as many as ten points may be energized. These points are wired directly to the patch board and are used to gate signals representing detected emitters to the system presentation. In addition, ten medium power sector activity switches and one switch for determination of the low power emitter environment are provided. The level selections are made by the quadrant relays.

The patch board and the special parameters unit content is variable. The patch board wiring and the output signal characteristics of the special parameter unit are determined by the requirements of the system utilising the simulator. The inputs required in order to make the system displays function realistically generally may be grouped into three classifications.

1. Digital word describing the characteristics of detected emitter.

2. Detected video waveform and associated analyzer outputs.

3. D-C indications to display and associated special threat signals.

A number of pre-programmed patch boards and special parameter units are used to fulfill the different system requirements. Figure 3 illustrates a patch board programmed to gate a video signal to the system display. The use of this patch board to distribute an input signal from the environment selection unit and gate video signals to the various system input channels is shown in Figure 4. In this instance, the special parameter unit is a video generator which will supply the required waveform combinations of representative pulse width and pulse repetition rates to the patch board for distribution.

DESIGN CONSIDERATIONS AND SYSTEM CAPABILITIES

The primary operator output of most reconnaissance and counter-measure systems is a visual presentation which indicates the states, activity and modes of operation of the system. Operator decisions are based on the intelligence presented by this display. The problem of simulation involves the realistic presentation of data at this display.

In order to overcome the disadvantages referred to previously in the introduction an electronic mapping technique will be used. The operational area is divided into quadrants which define fixed increments of latitude and longitude. This established a quadrant grid and each quadrant is uniquely identified with the RF emitters present in a particular geographic sector. Active quadrant group selection is determined by the air vehicle position. A finer rate of environment change may be achieved by using a multiple commutator with the segment groups out of phase. In the region near the pole, each quadrant relay is connected to several longitude lines.

The emitter content of each quadrant has been divided into three classifications with the respect to power level. A maximum detection range for each quadrant type has been determined. These ranges were determined by solving the beacon equation using typical power levels, antenna gains and a moderate receiver sensitivity. The detection capability is considerably beyond the line of sight range of conventional radars even when the air vehicle is at a maximum altitude where the line of sight is greatest. Taking these factors into consideration, the distribution of radars into power level categories will be made using range as a criteria. For purposes of this paper, consider radars having a power level capable of achieving a range in excess of 100 nautical miles to be included in the high power quadrant. The medium power quadrant will include all emitters having ranges greater than 20 nautical miles and all electromagnetic emission and disturbances having a lesser range will be included in the lower power quadrant.

The detection range of the air vehicle receivers can be much greater than the radar ranges discussed above, however for discussion purposes, the following detection criteria will be used:

Quadrant	Detection Range
High Power	Line of sight approx. 320 nautical miles
Medium Power	250 nautical miles
Low Power	90 nautical miles

The ranges are not based on the most optimum reception conditions.

An example of quadrant group selection is shown in Figure 5b. This represents a section of the position matrix. All latitude-longitude intersections in the matrix do not have an associated quadrant relay. Quadrant relays are located only in areas where an electromagnetic environment exists. The position of the commutators will select groups of quadrant relays by energising five latitude lines and five longitude lines. Figure 5a illustrates the degree of activation of the quadrant power levels. The air vehicle is operating in quadrant 1 and 24 associated quadrants are shown. Only the high power selection contacts of the outer ring of quadrant relays are energized and the middle ring of relays will have the high and medium power contacts energized. Quadrant 1 will have the selection contacts for all three power levels

energised. The circuit connections for quadrant relay 1 are shown in Figure 2. The position of the commutator allows all three selection contacts to be energised. These contacts thus determine the position of one High Power environment sector activity switch, one Medium Power environment switch and one Low Power environment switch position. This selection will activate a number one-wired points on the patch board. Similar selections will be made by each of the other 24 quadrant relays.

Variations in altitude will affect the extent and distribution of the air vehicle environment. To achieve a realistic change in environment with changes in altitude, an altitude variation selector switch will be used. The purpose of this switch will be to control the power distribution to the sector activity switches. Figure 6 illustrates the circuit connections for the altitude level selector switch. With each change in level the diodes either block or energise a new wafer of the environment selection switches. Assuming a ten wafer switch, approximately 25 to 30 invironment reduction steps will be available with change in altitude. To achieve a more continuous variation of environment with altitude, twenty or more levels may be used on the altitude switch.

The effect of antenna directional characteristics upon the activity presentation must be taken into account. Not only air vehicle coordinates, but the azimuth of approach as well as spontaneous manuevers will alter this presentation when signal strength is dependent upon directional criteria. If the foregoing effects are to be achieved, two parameters must be continuously available within the simulator system. These are: 1) air vehicle heading, 2) bearing of the emitter in question.

Since it has been hypothesized that system operation shall be a function of the coordinate analogs only, both of these parameters must be obtained on the basis of raw navigational data as presented by other subsystem simulators. For purposes of discussion, it will be assumed that these analogs take the form of coordinate shaft positions, although electrical analog or digital data forms could just as easily be processed through the use of an intermediate converter.

AIR VEHICLE HEADING

The directions of advance of both the x (longitude) and y (latitude) coordinate shafts are sensed by a differential heading encoder. Mechanically coupled differential units (see Fig. 7a) are used to sense the sign of each coordinate derivative by means of slip ring contacts. This establishes heading to within a quadrant since each derivative pair defines a two bit code as follows:

$\dot{x} > 0, \quad \dot{y} > 0$: 11

$\dot{x} > 0, \quad \dot{y} < 0$: 10

$\dot{x} < 0, \quad \dot{y} > 0$: 01

$\dot{x} < 0, \quad \dot{y} < 0$: 00

Heading is further defined (to within 45° of azimuth) by utilizing the additive and subtractive outputs of these differentials as the source of a third bit of information. If x and y are of opposite sign (second and fourth quadrants), the additive output is used, whereas if they are of the same sign (first and third quadrants), the subtractive output is used. The value of C is determined as follows:

quadrant codes 01 or 10:

$\dot{x} + \dot{y} > 0, \quad C = 1$

$\dot{x} + \dot{y} < 0, \quad C = 0$

quadrant codes 00 or 11:

$\dot{x} - \dot{y} > 0, \quad C = 1$

$\dot{x} - \dot{y} < 0, \quad C = 0$

The resulting three bit code defines one of eight possible heading sectors as indicated in Fig. 7b.

Since only the signs and relative magnitudes of the quadrature components of velocity are needed, the coordinate analog rotations may be augmented by some constant factor to improve resolution. Multiplication of this type (such as may be implemented by means of transfer gearing) will tend to establish a definite sense for either component even though variations in latitude and longitude may occur very slowly.

All of the functions described in the foregoing paragraphs could of course be accomplished by analogous electronic circuitry. The method discussed is intended to be taken as a first approach to the problem of generating heading information solely on the basis of slowly varying navigational data.

BEARING OF EMITTER

In order to provide detailed directional information on each radar a directional code will be included with each radar indication. An example of the code application is shown in Figure 8. The three bit code groups will be injected on each of the latitude and longitude lines. The combination of the two code groups will form a six bit code describing the relative position of the quadrant with respect to the quadrant containing the air vehicle. The five segments of each of the two commutators shown

in Figure 2 will transmit the code groups. A further interpretation of the code words will be made when they are correlated to the air vehicle heading. The resultant output will be a four bit binary description relating a radar source to the direction of detection relative to the air vehicle heading.

USE OF RANGE DEPENDENT GATES

An alternate approach to the quadrant selection method, and one which would be especially suited to the presentation of signal parameters in a digital format, would involve the use of resistive meshes and electronic switching. Under this method air vehicle coordinates at any instant define the origin of a range coordinate system such that the selection of geographical areas is completely dependent upon the mission flight path. Variations in line-of-sight range and emitter power relative to a given area are implemented as follows:

Assuming once more that air vehicle position is obtained from an external independent source in the form of analog shaft rotations, the instantaneous latitude and longitude may be represented by the output pairs of two coordinate distributors, each of which provide five range levels (See Fig. 9a). The highest level, designated as 8K, defines the immediate locus of the air vehicle to within one degree of variation, i.e. either x or y, while adjacent points receive levels of 2K and K respectively. When both coordinate contributions are summed, it may be seen that the three highest orders of range level are 6K, 5K and 4K. Thus, a point receiving a level of 6K is designated as being within the first domain of the air vehicle (or at a minimum range value) while the levels 5K and 4K define the second and third domains respectively. The range loci around a point corresponding to this order are shown in Fig. 9b.

It is evident from the figure that these loci as obtained on the basis of rectangular coordinates deviate considerably from the ideal circle locus of constant range. However, only the most elementary selections of range levels have so far been suggested. Additional combinations resulting from an increase in the number of levels generated will tend to improve the shape considerably.

Transistor gate circuitry may be used to sense the range information that has been established and then combine it with a quantization of emitter power to determine the final presentation of a given radar to the simulated activity display. A single gate unit may designate several radars of any given type at different locations provided that they are separated by at least three units in each coordinate. For an explanation of this behavior refer to the circuit of Fig. 9c.

During the "on" or saturated condition, the gate will both signify the presence of an emitter of predetermined characteristics and will pass coded information describing these characteristics to the processing centers. Saturation is determined as the sum effect of emitter power impedances in the base circuit and the range levels appearing at the base inputs.

Assume that the critical saturation current for a given collector load is I_1. Then a given emitter at (x_0, y_0) will go to the "on" state when the contribution of range levels from x_0 and y_0 establish a current $I \geq I_1$ through the base impedances. The expressions for I and I_1 are:

$$I = \frac{Vx + Vy}{R_p} \quad ; \quad I_1 = \frac{2K}{R_0}$$

If, for example, an emitter of characteristic power $2R_0$ is to be encountered at (x_1, y_1), I is given by:

$$I = \frac{Vx + Vy}{2R_0} = \frac{(m + n)K}{2R_0}$$

then the condition on range locus is:

$$m + n \geq \omega$$

or more generally

$$\frac{m + n}{k} \geq 1 \quad ; \quad \text{where } R_p = kR_0$$

The foregoing example illustrates the general approach to defining the criteria for signal reception subject to restrictions imposed by geographical separation and emitter power. An immediate objection to this type of mechanization would of course lie in the fact that different transistor gates will exhibit varying sensitivities because of variations in B and I_{co}. However, it should be borne in mind that, due to the nature of the problem, exact threshold levels need not be a prime requirement when, in fact, a gradual transition from the cutoff to on condition actually tends to approach real detection behavior even more closely.

SUMMARY

The simulator is composed of three types of functional assemblies, directional, environment selection and input-output modules. The directional and environment selection assemblies constitute the basic simulator and contain information on the distribution, intensity and bearing of all sources of RF emission within the operational area of the detection system. The combination of the basic simulator and selected input-output units enables the unit to be adapted to a large variety of systems which required a simulated environment.

The only input required is the position of the air vehicle. This input may be derived from the flight simulator in integrated training operations or from a mission programmer. The environment generated will provide a detailed and realistic presentation at the system display. The presentation may be made a function of heading and altitude. Variations in the environment are made by changing patch-boards. Test programs may be used which will enable the simulator to check out either itself or the video-digital sections of the detection system. Inherent advantages of the technique include:

a. Environment updating accomplished patchboard modification.

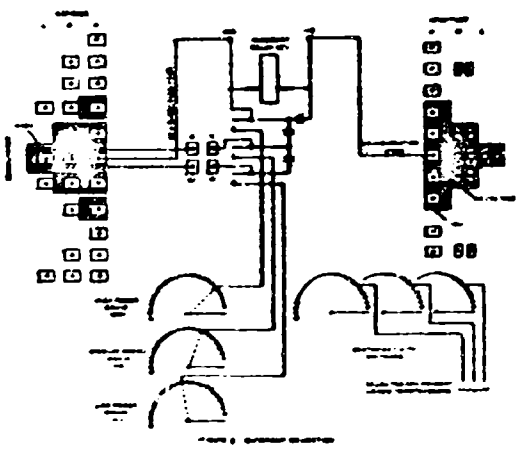
b. No restrictions on mission changes.

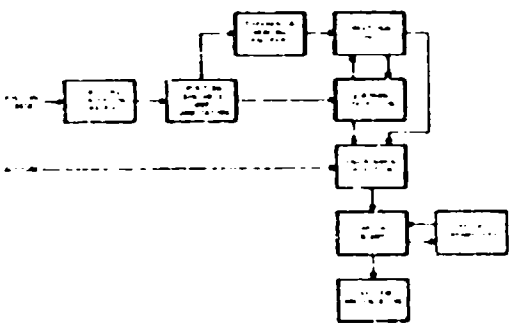
c. Patchboard may be changed to establish a new operating area.

d. Positional data is the only input required for integrated operation.

e. Provision is made for changes of coordinate system.

A simulator of this nature is somewhat analogous to modular automatic checkout equipment since it is compact, easily modified and requires no large expenditures of development time and money.





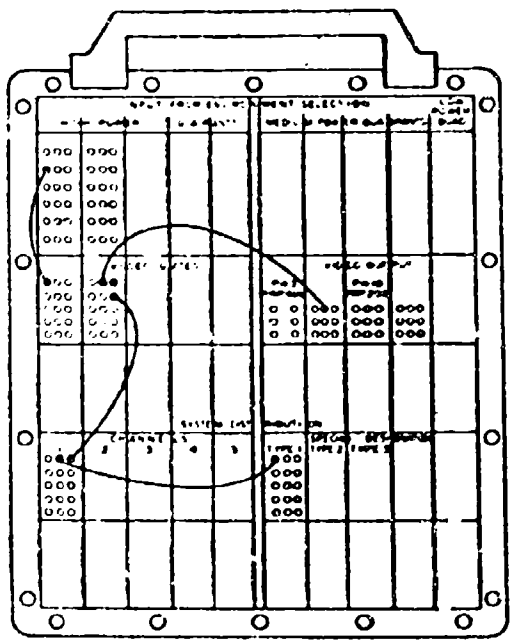
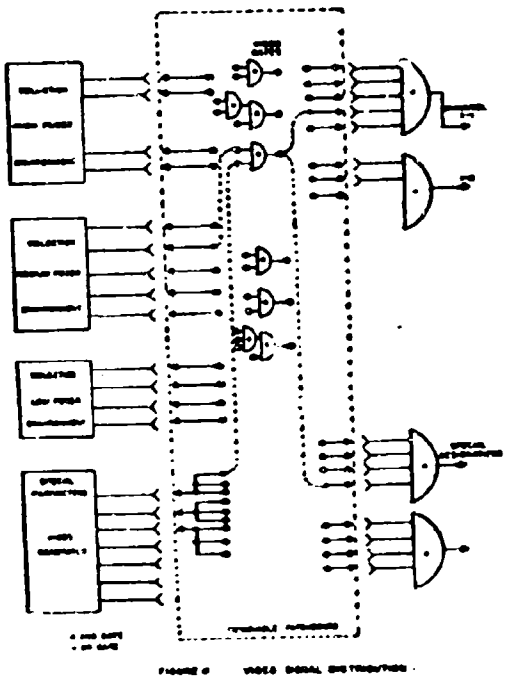
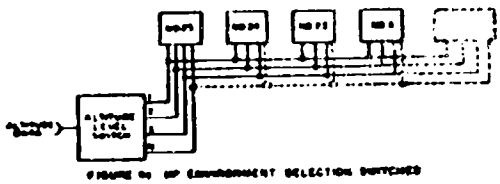


FIGURE 3 PROGRAM PATCHBOARD





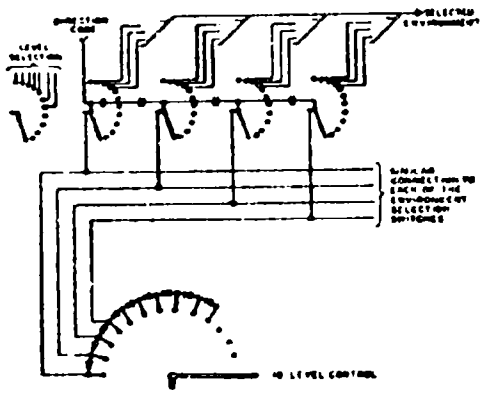


FIGURE 5b ALTITUDE LEVEL SWITCH

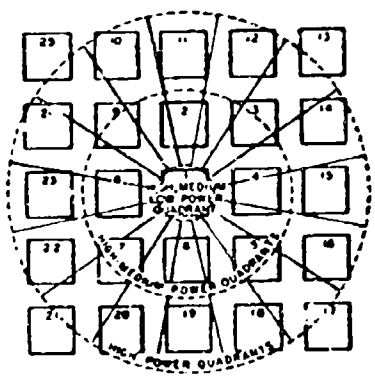
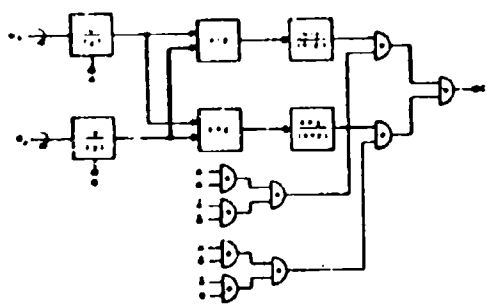


FIGURE 5a



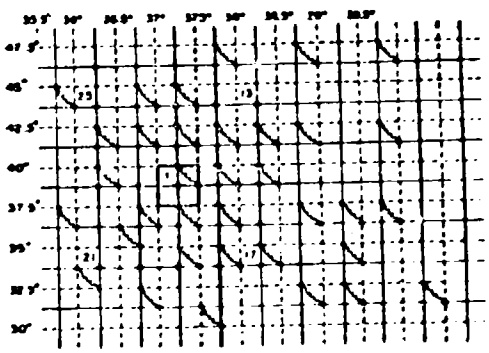
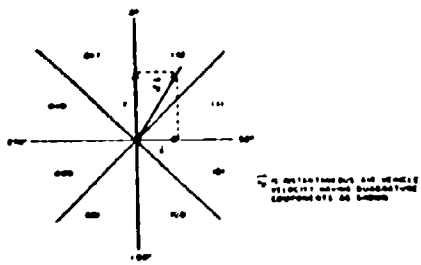
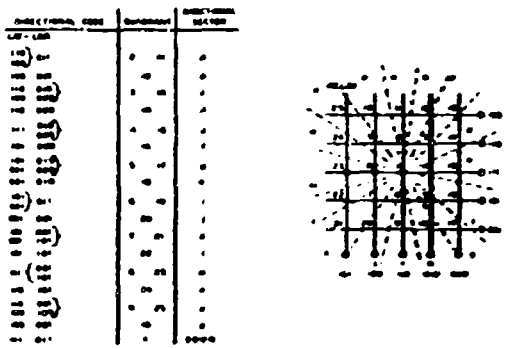
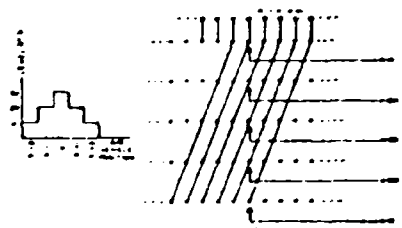
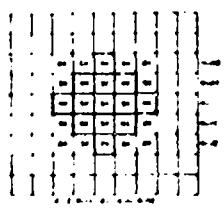


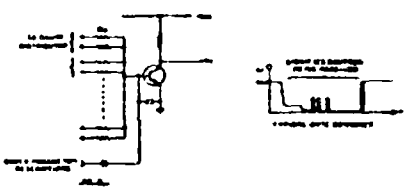
FIGURE 5b QUADRANT SELECTION











THE DEVELOPMENT OF A DYNAMIC TARGET AND COUNTERMEASURES SIMULATOR

By: Mr. R. L. Norton, MEWD, U. S. Army Signal Missile Support Agency

The expense involved in training radar operators to cope with raids involving hundreds of aircraft has led to the development of various target simulators.

As the state-of-the-art in electronic countermeasures (ECM) improved, it became apparent that ECM simulation was also necessary to train radar operators under the conditions that would prevail when aircraft would be radiating thousands of watts of jamming power. Most of the simulators developed used IF or video injection of target signals in the simulator and did not provide the realism desired or allow the radar operator to exercise electronic counter-countermeasures (ECCM) techniques effectively.

The problem then became one of realistic simulation of both target and jamming signals. Field experience at White Sands Missile Range (WSMR) had demonstrated that simulation could be accomplished most effectively at the RF level. This finding, by WSMR engineers, led to the design and development of the Dynamic Target and Countermeasures Simulator.

About two years ago a group of engineers and technicians at the Signal Missile Support Agency Laboratory at WSMR started development of this simulator. The original model was considerably different from the present developmental model since it was breadboarded from spare parts of salvaged radar sets and standard RF test equipment.

Operation was accomplished by receiving the transmitted pulse from the radar, time delaying the pulse to simulate range, mixing it with an appropriate jamming signal, and transmitting the target and jamming signal back to the radar. Tests were conducted on a selected group of radar operators and the excellent results obtained indicated that further development was warranted.

After this technique was proven in the field by the Signal Missile Support Agency, it was turned over to Gilfillan Brothers, Inc., Los Angeles, California, for further development and, finally, the fabrication of a mobile simulator unit. The present simulator is installed in a standard M-109 van which is a 12 by 7.5 foot inclosure mounted on a 2 1/2 ton truck chassis. The equipment provides for operation with L-, S-, and X-band radars.

As a training device, the simulator provides realistic target information in an ECM environment to such a degree that the radar system operators cannot distinguish any difference between a real or simulated target. Comparison of operator proficiency can also be made by programming the same simulated aircraft course for different operators. In addition to its use for training, it also may be used as a radar system checkout device. Signal saturation level can be determined and the use of ECCM techniques, such as regulating RF gain and video, contrast, brightness, persistence, and luminescence, can be checked.

GENERAL DESCRIPTION

The complete simulator consists of a control console, PPI monitor scope, computer, RF assemblies, power supplies, and such minor subassemblies and couplings as are required to connect the equipment into the radar system. Figure 1 shows part of this equipment. The PPI monitor scope, located above the control console, permits the simulator operator to monitor the operation of the radar L-band and S-band PPI scopes. Simulator video may also be observed.

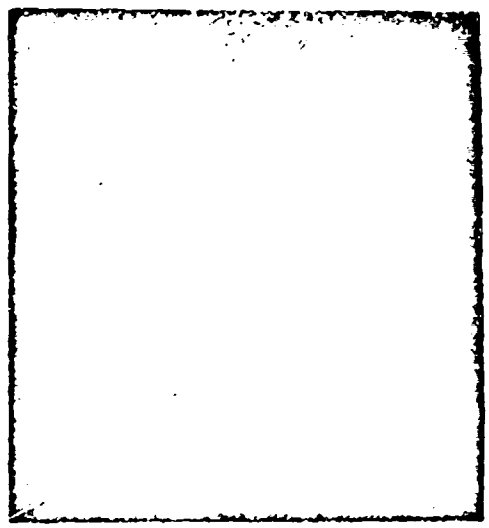


Figure 1. Control Console.

Basically, the simulator is used in conjunction with a ground-based acquisition radar and target track radar. The acquisition radar is supplied with a simulated target and jamming information through a directional coupler at the antenna. This allows the radar to operate normally, and since the antenna is revolving, ground clutter and actual moving targets will be seen on the PPI scope with the simulated target added.

The target track radar receives its simulated data through a horn antenna mounted on the radar systems RF test mast. When the simulated target is designated by the acquisition operator, the target track antenna slews to this beacon on the test mast. The target track antenna will "lock on" to this beacon. Error voltages which provide for apparent motion of the antenna are then supplied by the simulator. In addition, circuitry is provided to cause the target track radar azimuth and elevation indicator dials to continuously register the correct

simulated target position. Operators will track normally and not be aware that the antenna is fixed in azimuth and elevation. During actual operations, the acquisition radar operator electronically designates any target on his scope and it will be transferred automatically to the target track radar. Therefore, the operator may designate either a real or the simulated target and it will be transferred to the target track radar. If the real target is designated, logic circuitry in the simulator disconnects the simulator from the target track radar which allows the antenna to slew to the real target and track normally.

If either a simulated or a real target is transferred, the time from target designation by the acquisition operator to the time of target acquisition by the target track radar operator can be measured for various operators to determine their proficiency.

TABLE 1

Target and Jamming Source Parameters

Speed	0-2800 knots
Altitude	0-100,000 ft
Azimuth	0-6400 mils
Turn Rate	0-6 deg/sec
Maximum Range	150 NM
Heading	0-360 deg
Jamming Source Power	1-100 w/mc
Target Radar Cross Section	0.1-100 m^2

Target scintillation and range attenuation ($1/R^2$ for the jamming source signal and $1/R^4$ for the target signal) is also provided. It is to be noted that only one target and one jamming source is provided in this developmental model. Both the target and jamming signal are independently "flyable" in heading, range, altitude, and speed. The simulator operator merely sets any initial courses desired and then programs these within the parameters already mentioned. Although one target and one jamming source is provided, the addition of any number of simulated targets and jamming sources is possible with very little increase in space requirements.

CONTROL CONSOLE OPERATION

The control console (Figure 1) has controls with which the operator can set in the initial and flight conditions of the target and jammer. Initial conditions consist of the initial azimuth, altitude, and ground range of the target and jammer. Flight conditions are speed, turn rate, and heading of the target and jammer. Target area and jammer power controls are also provided. As a further aid to the simulator operator, dial indicators on the computer units, immediately above the control console, display the present range, elevation, azimuth, and heading of both the target and jammer. The PPI monitor scope, as previously explained, allows the simulator operator to monitor the acquisition radar PPI scopes as well as simulator video information.

COMPUTER OPERATION

The computers for the target and jamming source data generation are identical; therefore, the target computer will be explained as a matter of convenience. The computer is a combined static and dynamic coordinate-data generator and the function of each section is outlined in Figure 2. Initial conditions are set into the course generator when the simulator operator positions the appropriate control knobs at the control console. Initial ground range (R_g) setting, for example, provides analog voltages proportional to the starting position of the target in ground range. These reference voltages are applied to the input of a sine-cosine potentiometer. Selection of the initial target azimuth then positions this potentiometer by means of a servo mechanism. The output of the potentiometer becomes R_g Sin θ and R_g Cos θ or, as shown in Figure 2, $X_{T(I)}$ and $Y_{T(I)}$, the initial rectangular coordinates of the target. In the static mode, at any selected altitude, the simulated target can be initially located at a suitable range off scope, and programmed to "fly" into the field of view from any desired location.

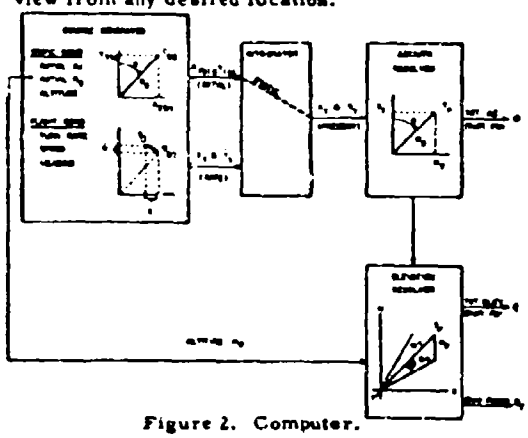


Figure 2. Computer.

Flight conditions which include target speed, heading, and turn rate are then set into the computer (Figure 2). A similar sine-cosine potentiometer arrangement is used with the speed control providing analog voltages as inputs which are proportional to target speed. Target heading and turn rate then positions the potentiometer shaft through a servo mechanism. The output is the $\dot{X}$ and $\dot{Y}$ rate information.

When the simulator operator positions the Target Status Switch at "fly," $X_{T(I)}$ and $Y_{T(I)}$ (initial position of the target information), and $\dot{X}$ and $\dot{Y}$ (the rate information) are integrated. Rectangular coordinate data indicating the present position of the target, X_T and Y_T, is obtained.

The azimuth resolver section converts the rectangular coordinates of the target into polar coordinates, these being the present ground range (R_g) and the target azimuth (θ). Target azimuth is represented by the shaft position of the potentiometer and, as explained later, is used in azimuth gating of the acquisition radar target and jamming signals.

Present ground range (R_g) is supplied to the elevation resolver section as shown in Figure 2. This information, along with initial altitude data, set into the computer, provides the polar coordinate, slant range (D_T), and the angle of elevation of the target (β). Slant range in the form of a DC analog voltage is then supplied to the precision range unit.

OPERATION OF RANGE UNIT

The precision range unit makes use of the Meacham system, which utilizes a quadrature phase shift capacitor, to provide a jitter-free pulse delay generator. Trigger pulse (preknock) is obtained from the radar system and supplied to a timing wave generator and a delay gate generator. Range delay in relation to the preknock is then obtained, based on the target or jammer slant range information from the computer. Pulses representing the proper range delay are transmitted through coaxial cables to the appropriate RF signal generator in the RF assemblies.

OPERATION OF RF ASSEMBLIES

L-, S-, and X-band RF assemblies have been provided with this equipment. Since the L- and S-band assemblies are similar in operation, only the S- and X-band units will be discussed.

Figure 3 illustrates the method used in providing RF target and jamming information. A considerable number of "off-the-shelf" items have been used in fabricating this simulator, including 13 traveling-wave-tube (TWT) amplifiers. Here, the TWT amplifiers are cascaded in both the target and jammer chains. Initially, the range pulse from the precision range unit triggers the S-band signal generator to produce the RF target signal which is then amplified by TWT Number 1 and TWT Number 2. Both of these TWT's are modulated by operational amplifiers (OA in Figure 3) with inputs from an appropriate network. The input provides for free space attenuation of the target signal, radar target cross sectional area control (0.1 to 100 square meters), target scintillation, and antenna pattern simulation. Target range is attenuated inversely with the square of the range by TWT Number 1 and the output again similarly attenuated by TWT Number 2. The target signal, provided as input to TWT Number 5 has, therefore, been attenuated inversely to the fourth power of the simulated target range; TWT Number 5 provides signal amplification to compensate for attenuation in the coaxial cable to the acquisition antenna.

The jammer chain in Figure 3 uses direct internal noise amplification of TWT Number 3 to produce broadband "white" noise. One hundred watts per megacycle jamming power can be obtained at a simulated range of three nautical miles. Here again, as in the target chain, variables such as jamming power (1-100 w/mc), antenna pattern effects, and $1/R^2$ attenuation for the jammer range are summed and applied as modulation control for TWT Number 4. The particular technique of antenna pattern reproduction provides realistic jamming presentation on the radar PPI scopes as the jamming power may enter the side and back lobes of the antennae.

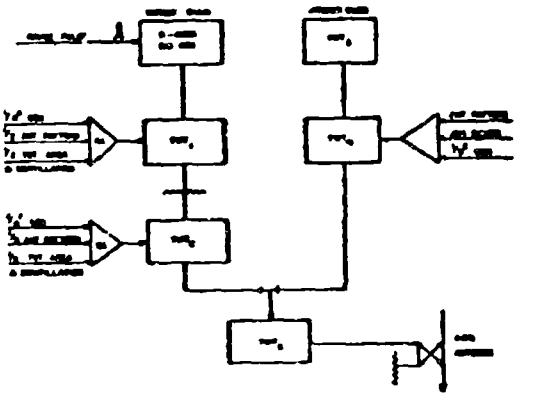


Figure 3.
RF Assembly, Acquisition Radar.

The RF assembly for the target track radar is slightly different from the acquisition radar arrangements, as shown in Figure 4. As previously pointed out, the simulated target and jammer signals are transmitted from a fixed beacon. Antenna pattern simulation is not necessary since the RF signals are transmitted through space to the actual antenna. During operation with range pulse from the computer triggers a signal generator to produce the RF signal which then passes through a ferrite modulator to provide target scintillation. Space attenuation is obtained by the use of two variable waveguide attenuators. The waveguide attenuators are positioned by a control motor which is operated by an amplified voltage inversely proportional to the square of the range, thereby automatically setting each attenuator to provide $1/R^2$ attenuation. Proper target or jamming source signal attenuation is then secured by passing the signals through one or both attenuators as appropriate. Jammer chain operation, it will be noted, is similar to that employed with the acquisition radar.

As a point of interest, the $1/R^2$ voltage utilized for signal attenuation effects is obtained by supplying the computed target or jamming source slant range voltage to a diode function generator. The output of this function generator is proportional to $1/R^2$ and is one input to the TWT driver amplifiers.

Target area and scintillation is provided through the use of a Rayleigh noise generator. The noise generator receives its input from the target area control located on the control console. This control allows the simulator operator to select any cross sectional area from 0.1 to 100 square

meters to represent a specific missile or aircraft. The filtered output also provides the target scintillation effects. Here again, the output is used to operate the TWT driver amplifiers for TWT modulation.

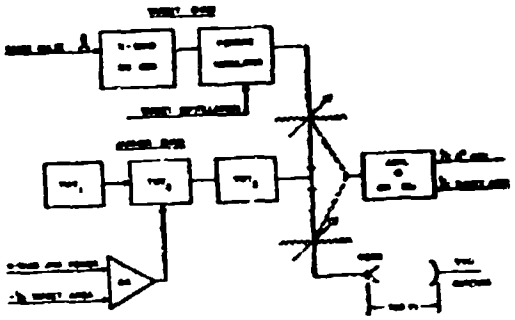


Figure 4.
RF Assembly, Target Track Radar.

ANTENNA PATTERN GENERATOR

The antenna pattern generator (Figure 5) developed for use with this simulator provides realistic reproduction of the antenna patterns of the L- and S-band radar antennas. A variable area photographic film is made of the actual antenna pattern. The film is then inserted in the generator unit and rotated in synchronism with the acquisition antenna. The main lobe on the film is offset an amount equal to the simulated target or jammer azimuth. A light source, projected through this rotating film, activates a photocell to produce a voltage inversely proportional to the antenna pattern. The amplified output is used to drive a TWT chain in the RF assembly. Excellent results have been obtained using this device since, as jamming power is increased, the jamming signal will enter the side and back lobes realistically.

Input information for the pattern generator is obtained from the resolver sweep circuit of the radar and from the simulator computer. Angular data relative to the acquisition antenna rotation is necessary if target information is to be presented at the proper azimuth on the radar PPI scope. This data is obtained from the radar resolver sweep circuit in such a manner that the antenna pattern generator shaft will rotate in synchronism with the acquisition antenna azimuth scan.

A mechanical differential is driven at the same rate of rotation as the acquisition antenna and the differential output shaft then drives the antenna pattern generator. Film is thereby positioned so that the simulated target signal will appear in proper orientation.

FUTURE DEVELOPMENT

This present developmental model is now undergoing test and evaluation by the Army. The outcome of these tests will determine the future utilization and refinement of this equipment. In designing this unit, cabinet space was allowed for future expansion which may include the addition of chaff generation equipment and an increased number of simulated targets and jamming sources. This could be the next step in the development of this simulator.

By slight modification of the control console, magnetic taped courses, in the form of standard operator electronic countermeasures tests, could be programmed and radar operator response recorded. Standard scoring on a comparative basis could be initiated to determine operator proficiency.

During the design and development of this simulator, every effort has been exerted to make target simulation and the jamming environment realistic. With the state-of-the-art in countermeasures progressing at such a rapid rate, modern radar target simulators must be designed to allow the radar operator complete use of all electronic counter-countermeasures techniques available to him. Further, when the operator applies a specific countermeasures technique, the results obtained must be exactly the same as those obtained with actual targets in a similar countermeasures environment.

By providing realistic target and countermeasures simulation, radar operators can be trained effectively on a round-the-clock basis, with a considerable saving in manpower and money when compared to the cost of actual aircraft training missions.

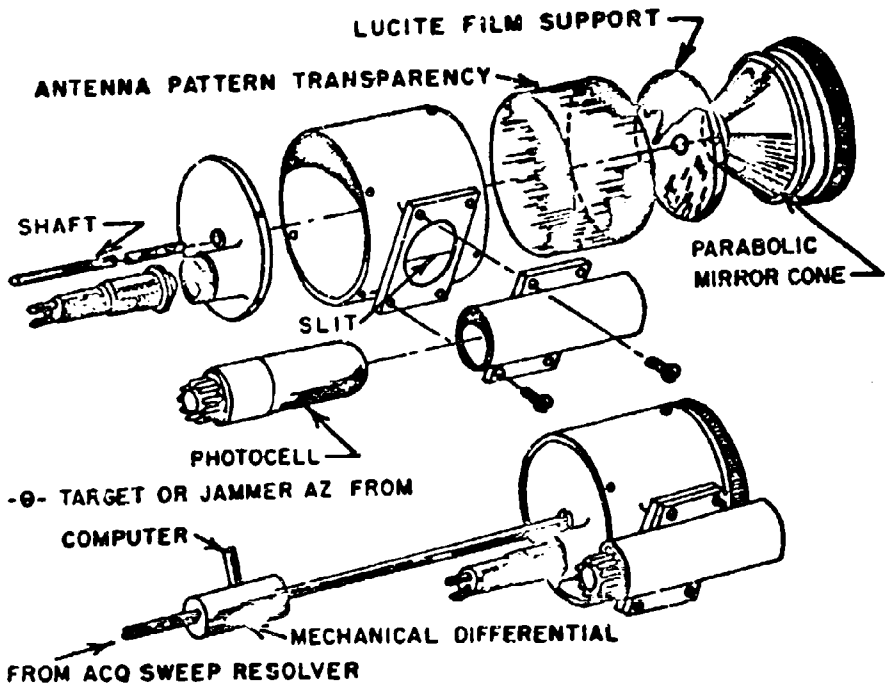


FIGURE 5, ANTENNA PATTERN GENERATOR

CELESTIAL NAVIGATION
By Gerard Joquies

A. INTRODUCTION

Before the invention of Hadley's sextant in 1731, little progress had been made in celestial navigation. With the publication of the principle of the intercept method of Marcq. St. Hilaire in 1875, the practice of celestial navigation began to take its present shape. Advances during the past seventy-five years, important though they are, should be regarded more as a refinement of technique rather than new fundamental principles.

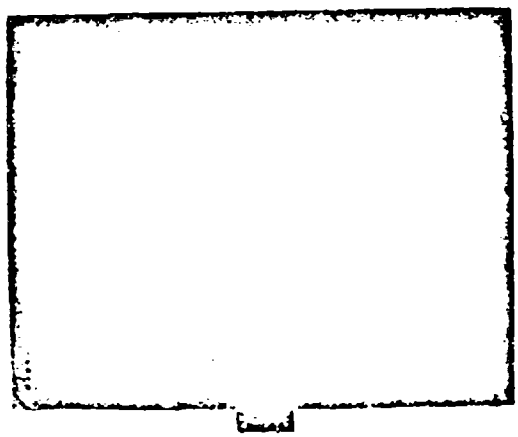
Early navigators were trained by the "apprenticeship" method. This method proved to be a matter of chance rather than a method of selection, since the choice of an apprentice was usually left to the "master". Once chosen, there followed years of travel and training. The instruction was of an individual nature and in manycases perpetuated the errors of specific ideas of the master under which one served. In addition, if most of an apprenticeship was spent in a particular geographical area, difficulty was later encountered in other global areas due to the inexact knowledge of the relationship of the different points on the earth's surface to the celestial sphere.

The advent of aircraft, World War II, and the development of certain types of missiles has accelerated the need for trained navigators, therefore, the requirement for navigational training devices.

The concept of training devices is not new, as the history of the art indicates. Early devices consisted of such individual items as the armillary sphere, the Astrolabe, and the terrel. These items were used in the training of mariners, and pre-date the quadrant, sextant, and compass by many years. They were, of course, the objects used in "individual" training programs and did little to cut down the actual time required to train a navigator. Our most modern training devices are designed for individual and/or group training in the shortest possible time, under simulated conditions.

Just prior to World War II it became evident that individual training for thousands of navigators was impossible due to lack of planes and pilots, but of equal importance was the exorbitant cost even if they were available. During the war a series of celestial trainers was developed for the Navy beginning with Device 1A. In 1958 the final simulator of this class was built and designated as the 1A18A. This device consisted of an eighteen foot celestial dome, a student observation station, an instructor console, containing the computer and controls, and 5 student booths with their respective DR instruments. The dome contained 12 collimated navigational stars and was mounted on gimbals to permit motion in two directions, and with the capability of polar navigation over the North Pole. This series of trainers was housed in a building resembling a familiar farm structure 44 feet high and 22 feet wide, well known as a silo. (Fig. 2)

The Air Force redesigned this celestial dome class of trainer and designated it as the D-2 (Fig. 1)



The celestial dome of the D-2 is 45 feet in diameter with facilities for five students to take simultaneous sights. The sighting was accomplished with the periscopic sextant through a unique optical system. Economics has limited production to one unit. In order to obtain the required accuracy, the structures in which the dome type trainers are housed require constant temperature and humidity control. The shifting of this equipment, due to re-location of training commands was impractical.

B. TRAINING REQUIREMENTS

Let us now consider the requirements of a modern celestial navigation trainer. For maximum utility the device should provide for group (class-room) or individual training in celestial and dead reckoning navigation. It should be capable of covering the Northern and Southern hemispheres, including the Polar regions. In effect, the trainer should simulate the navigational conditions presented by modern high-speed aircraft, where the navigator is required to derive flight data from celestial observations and dead reckoning instruments anywhere on earth.

The trainer should be designed to teach the use of the operational sextant and sky compass in deriving sun and star information. Dead reckoning

information, such as barometric altitude, radar altitude, heading, free air temperature, air speed and time should be displayed to the students so that an accurate log and plot of the flight path can be kept. Additional data such as wind parameters and ground speed may be made available through the use of a driftmeter, or similar ancillary devices.

The dead reckoning instruments should be so located that they may be monitored continuously by a group of students. Readings on the sextant sky compass, and driftmeter occupy only a small portion of the individual navigator's time and and may be accomplished by one student at a time during the course of a problem.

Astro data should show correctly-positioned stars and/or sun. Accurate readings of these displays should be possible by using the periscopic sextant. A separate display simulating polarized sky light should be provided, correctly oriented as seen at the zenith, and read by means of the sky compass. A graphic ground track recorder should permit the instructor to continuously monitor the simulated flight path, which may then be compared with the logs of the individual students. The recorder display may be either Mercator or Polar plot, depending upon the region over which the "flight" takes place.

C. GENERAL DESCRIPTION

Last year a paper was presented by Dr. M.D. Ben Bennett and Mr. W.E. Mickelson of the Reflectone Electronics, Inc. on the design considerations for a Celestial Navigation Trainer. Since that time there has been an operational evaluation and several minor modifications. It is the intent of this discussion to delve primarily on the results of the evaluation, the accuracies obtained, and the proposed changes to satisfy future requirements. The detail technical function of each unit discussed in the previous section have not been repeated.

Prior to the design of any of the hardware for the device, the computer was blocked out and a rigorous mathematical analysis of the computer functions performed in order to determine the accuracy that might be expected from the completed trainer. The mathematical analysis revealed that under certain conditions of operation an ambiguity would develop in the angle of declination computer. The circuitry for this computer was revamped to avoid the ambiguity and the design of the computer was carried to completion. As predicted by the mathematical analysis, the accuracy of the completed trainer fell well within the specification requirements.

Among the several unique problems that developed in the design of this equipment, was the design and construction of the collimated light representing a star of the second magnitude as seen through the periscopic sextant. In order to avoid aberration that might result from the use of a photographically produced reticle, the final reticle was formed by drilling a hole .0008 inches in diameter through .001 hard rolled cadmium plated brass special blue tinted plastic screen was used between the light source and the reticle to obtain the characteristic blue color of the star. The front or cadmium plated face of the brass was masked and front illuminated to represent the sun. High order optics were required in precise alignment in order to permit viewing of the star through the periscopic sextant while allowing for the normal excursions of the entrance pupil of the sextant in its conventional mount. It is necessary, as in the operational problem, to hold the sextant bubble level.

Scaling problems were solved by and large through the use of precision wound toroids tapped on a decade basis.

All servo motors are driven by magnetic amplifiers. Thus, their operation is essentially maintenance-free.

The Celestial Navigation Trainer, figure 3 as manufactured by Reflectone Electronics, Inc. is a classroom trainer for training students singly or in a group in celestial and dead reckoning navigation in the northern or southern hemispheres, including the polar regions. The trainer simulates the navigational conditions presented to a navigator in modern high-speed aircraft wherein the navigator is required to derive flight data from celestial observations and dead reckoning instruments.

The trainer is designed to teach the use of the operational sextant and sky compass in deriving star and/or sun information. Also presented to the student is dead reckoning information as displayed in an aircraft, such as barometric altitude, radar altitude, heading, free air temperature, air speed and time, in order that the student may keep a correct log and plot the flight path. Additional information may be derived by means of a driftmeter simulator to yield data on ground speed and wind. The trainer also provides synchro signals for driving the movable transducer of a Loran trainer.

The dead reckoning instruments may be monitored continuously by a group of students. Readings on the sextant, sky compass, and driftmeter are accomplished by one student at a time during the course of a problem.

The astro information comprises a correctly positioned single spot of light simulating a star or, alternately, a source of light representing the sun. Readings of these displays are obtained by means of the periscopic sextant. The star display includes a selection of three predetermined stars, of which the student may choose one at a time for sighting.

A separate display simulates polarized sky light, correctly oriented as seen at the zenith. It is read by means of the sky compass.

A graphic ground track recorder permits an instructor to continuously monitor the simulated flight path which may be compared against the log of the individual students. The display is either Mercator or Polar plot, depending upon the region over which the "flight" takes place.

D. INPUTS

The trainer is operable within the problem area 0 to 90 degrees north or south latitude and 360 degrees in longitude. Individual problems are limited, in general, to the area covered by the recorder charts in use. The general characteristic and range of operation of the trainer are as follows:

a. Wind:

(1) Wind Speed	0 to 300 Knots
(2) Wind Speed Change Rate	0 to 50 Knots per hour
(3) Wind Direction	0° to 360°
(4) Wind Direction Change Rate	0° to 30° per hour

b. Air Speed:

(1) True Air Speed (two ranges)	0 to 300 or 0to 1500 Knots*
(2) Indicated Air Speed	0 to 250 Knots (Dial limited)

* In other models this range can be extended beyond 3200 knots.

c. Temperature:

(1) Free Air Temperature (aloft)	50° to -75°C.
(2) Ground Air Temperature	50° to -75°C.
(3) Lapse Rate	-2°C per thousand ft.

d. Altitude:

(1) Absolute Altitude	0 to 80,000 feet
(2) Vertical Speed	0 to 15,000 feet per minute
(3) True Altitude	0 to 80,000 feet
(4) Barometric Pressure	28.0 to 31.0 inches of mercury
(5) Pressure Change Rate	0 to +200 feet per hour

e. Heading:

(1) Magnetic Variation	180° East or 180° West
(2) Magnetic Deviation	0° to +19°
(3) Compass Heading	0° to 360°
(4) True Heading or Grid Heading	0 to 360°
(5) Gyro Error	+180°
(6) Gyro Precession Rate	0° to +15° per hour
(7) Left or Right Turn	3° per second

f. Astro Data coordinates of any celestial object as follows:

(1) Star Declination Angle	0° to 90° North or So.
(2) Sidereal Hour Angle	0° to 360°
(3) Local Hour Angle	0° to 360°

g. Time Indication:

(1) Civil Time (standard aircraft clock, master and slaves)	12 consecutive hours
(2) Problem Time (digital clock)	Elapsed time of problem (24 hours)

h. Recorder Charts (see table 1-2):

(1) Mercator Charts, all 25° wide in longitude:

Chart I	Minus 1° to plus 31° latitude N or S
Chart II	31° to 55° latitude N or S
Chart III	55° to 70° latitude N or S

(2) Polar Grid Chart with pole at center

Chart IV	70° to 90°

i. Ground Speed to Driftmeter: 700 knots maximum

j. Drift Angle to Driftmeter: 30°

k. Altitude Limits on Driftmeter operation: 3,000 to 30,000 feet

l. Ground Speed to Loran: 1500 knots maximum

m. Ground Track Angle to Loran: 0° to 360°

n. Star Brilliance: Second Magnitude

E - COMPONENTS

The trainer comprises four major units, consisting of the Instructor Console, Computer-Recorder Console, Student Observer's Station, and Classroom Instrument Panel, in addition to a separate Driftmeter Simulator. All of these units are movable and are interconnected by electrical cables.

The Instructor Console houses the necessary power distribution and controls, switches, etc., required to set up navigation problems. It also contains the controls for adjusting the input values for the star representations selected.

The Astrodata Panel is mounted on the upper center portion of the console. It incorporates three identical panels, each containing dials which control and indicate the introduction into the computers of quantities representing celestial coordinates of one star. Each panel contains a card holder for insertion of a card which bears identification of the star represented. Each panel provides a control dial (0° to 5°) for inserting a value of SIDEREAL HOUR ANGLE which appears on an indicating dial (0° to 360°), three dials for inserting the quantity of sine of the star declination angle, and three dials for inserting the quantity of cosine of the star declination angle. Each panel includes an indicator lamp which, when energized, indicates that the student in the observation station has selected that particular star for his navigational computation. A switch marked DECLINATION NORTH-SOUTH provides the proper hemispheric factor of declination.

The Computer-Recorder Console houses the necessary mechanism to compute the flight course as determined by the settings at the Instructor Console and to compute the position of the stars relative to the simulated aircraft (Student Observer's Station). A recorder mounted on the outside of the console provides an automatic ground plot either in Mercator or Polar projection of the course "flown" by the aircraft as represented by the Student Observer's Station. This plot may be used by the instructor to evaluate student performance by comparing the plot against the log of each student.

The Computer-Recorder Console further houses the computing equipment for providing the synchro signals (ground track angle and distance to a movable Ionna Transducer).

The Student Observer's Station consists of an enclosed booth with access door. Red illumination is provided within the booth for eye adaptation to night conditions. One wall of the booth is designated as "forward". The booth actually represents the aircraft, and the representations of star or sun and of polarized light are positioned relative to this forward wall of the booth.

One area of the booth serves for use of the sky compass which is used to look at a simulated sky. The sky comprises a polarizing plate which is illuminated from a light source, thus providing polarized light. The computing mechanism drives the plate continually in azimuth so that the instantaneous position of the plate yields polarized light to the sky compass from the proper azimuth angle. The intensity of the light source is controlled from the Instructor Console.

The other area of the booth is used for simulating the star or sun representation. An overhead movable star-sun mechanism contains sources of light which represents the stars or sun and which are viewed by means of the periscopic sextant. This mechanism mechanically moves in azimuth and elevation. The computing mechanism of the trainer continuously provides azimuth and elevation position information to the star-sun mechanism. When the student operates the selector switch provided in the booth, the star mechanism assumes the correct coordinates of the star (which have been selected relative to the forward wall of the booth) and a sighting may be taken. Subsequently, upon operation of another selector switch, sighting of one of the other star positions may be taken. The source light representing the star is projected through a pin point aperture while the sun representing the star is projected through a pin point aperture while the sun representation consists of an illuminated lucite disc, both displays being viewed with the sextant through a collimating arrangement within its mechanism.

The Classroom Instrument Panel is a flat, rectangular, portable panel mounted in view of the group of students. The panel contains six instruments, 19-inch in diameter each, which simulate the following operational instruments: Barometric Altimeter, Radar Altimeter, G-2 Compass, Free Air Temperature Indicator, Air Speed Indicator, and Time Clock. These instruments are continuously monitored by the student body. The instruments are actuated by synchros which receive their signals from respective synchros at the Instructor Console.

The Barometric Altimeter is provided with an adjustment so that students can correct for the barometric pressures inserted by the instructor. The G-2 Compass is provided with indicator lights which indicate its conditions of operation. The Air Speed Indicator is provided with indicator lights which are actuated from the Instructor Console to inform the students whether the instrument is indicating true air speed (TAS) or indicated air speed (IAS). The Air Speed Indicator simulates the conventional Maximum Allowable Air Speed Indicator. The maximum pointer and Mach scale are inoperative.

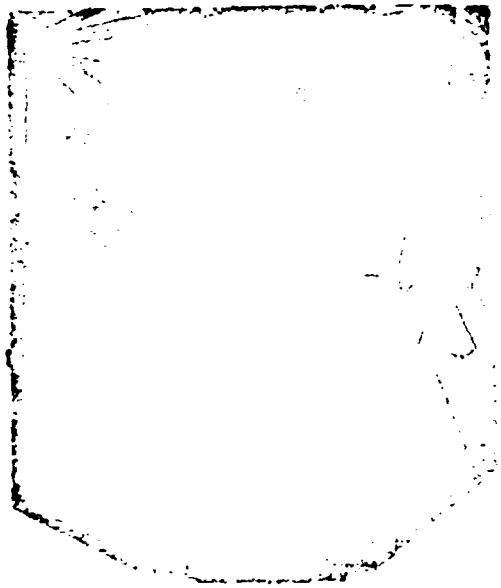
F - MODIFICATIONS

Upon installation of the equipment at Barber's Point, it was found that certain mechanical difficulties had developed in the local

hour angle (LHA) computer and that the damping of the collimated star drive was marginal. A review of the mathematical analysis revealed that insertion of single step solution of the sine and cos of an angle did not contribute to a stable solution. Consequently a two step or tangent solution was introduced into the resolvers simultaneously which resulted in greater accuracy and stability. An example of the change is as follows:

a. Astro Data Panel

The Astro Data Panel (figure 4) enables the instructor to insert the celestial coordinates of any celestial body. The sine and cosine of the declination angle are separately inserted by the "Sine Declination" and "Cosine Declination" controls which operate induction type voltage dividers. The amplitude of these output voltages are proportional to sin d and cos d. Sin d is fed to the Computer Recorder Console and cos d is fed to the Latitude Computer. The sidereal hour angle (S.H.A.) is inserted by the "Sidereal Hour Angle" control which positions synchro transmitters 1FB1 and 1FB2. The S.H.A. is fed to the Computer-Recorder Console.



Before the device was modified cos d was sent directly to the Local Hour Angle Computer (L.H.A.) (Figure 5) as generated by a reference voltage from the cosine declination voltage divider in the Astro Data Panel. This reference voltage cos d data is now fed to the Latitude Computer in the Recorder Console Section. Cos d proportioned voltage is combined with the rotation proportional to latitude by resolver 2DB5, resulting in an output voltage the amplitude of which is proportioned to cos l cos d. This voltage is then returned to the L.H.A. panel and combined with a rotation proportioned to L.H.A. by resolver 1BB5 resulting in out put voltages whose amplitude are proportioned to cos l cos d cos t and cos l cos d sin (-t).

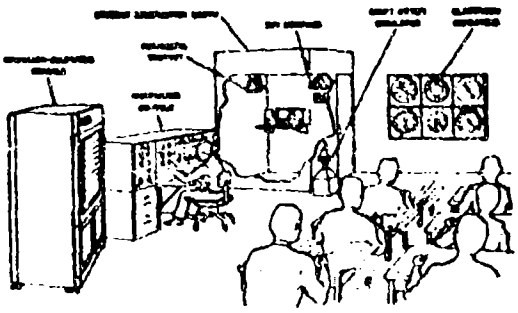


FIGURE. - 3

This latter value is fed to the Star Azimuth Computer (Figure 6), the output of which positions the Star in Azimuth. The above changes, with some associated damping on the servo motors outputs, changed a marginal system into one which is remarkably stable and maintainence free after hours of continuous use.

C- EVALUATION AND FUTURE REQUIREMENTS

Upon completion of these changes a formal evaluation of the equipment was performed with the following results:

Input	Accuracy
Sextant altitudes	± 5' minutes
Plotted fixes	± 5 miles
Track	± 1° (degree)
Time	no error
Azimuth Indications	1° (degree)
Speed	5 knots
Final results	98½ to 99% accurate

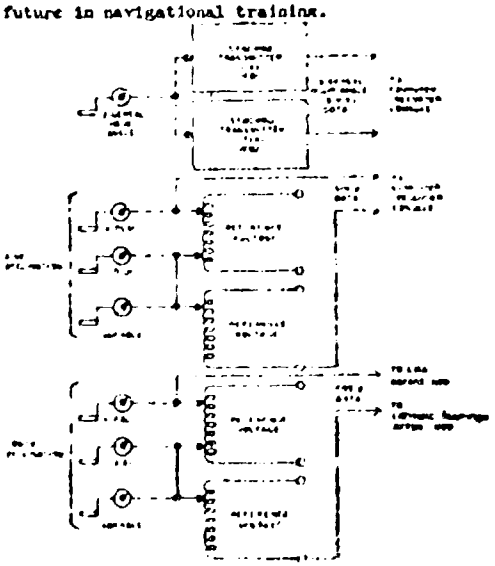
Overall effectiveness indicated that a problem can be set in the simulator and readied for a class in less than 30 minutes. The ability to stop or freeze the problem for class discussion and restart without loss of accuracy is extremely desirable. Also the ability to simulate polar flights over either pole has not been possible in the past. A Squadron Navigation Instructor commented that, "the 1A19 tracked perfectly in all circumstances and simulated a flight in an excellent manner. The drifts derived from track and ground speed and pressure pattern were identical in all cases. The

students adapted well to the trainer and they derived an immense amount of training through its use."

The evaluating activities recommend that Loran facilities be adapted to the simulator for more thorough training in DR navigation. It would be desirable to have a navigation facility with 2 or more sextants operating from the same computer so that additional students may take simultaneous fixes. This recommendation has been under consideration and steps are being taken to adopt this change in future productions.

The results of the evaluation of the Celestial Navigation Trainer, Device 1A19, indicate that it is possible to provide for rapid Training of navigational personnel under any and all geographical conditions. The device is in no way dependent upon climate, atmospheric conditions, or other limiting local conditions. It is based upon operational equipment and provides for optimum training of groups or individuals. Its accuracy is well within the limits required for practical navigation. Finally, it requires no complex operating instructions or highly-complicated charts for operation.

In this age of space exploration when ballistic missiles and jet speeds of Mach 2-3 become common, high-accuracy navigation is essential. A recent paper presented to the IRE on coordination for high-accuracy navigation by Capt. A.B. Moody, indicated a possible solution is possible by adopting latitude and longitude to a universal coordinate system. The importance of the training in the technique of obtaining latitude and longitude position by celestial means is extremely essential. With this in mind, the technique as utilized in this celestial simulator will serve the needs of the present and the future in navigational training.



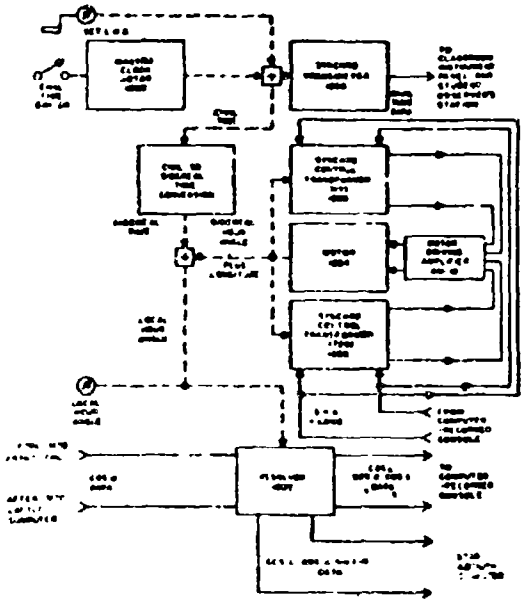


Figure 5 View of I.P.I. Panel (Instructor's Console), Functional Diagram

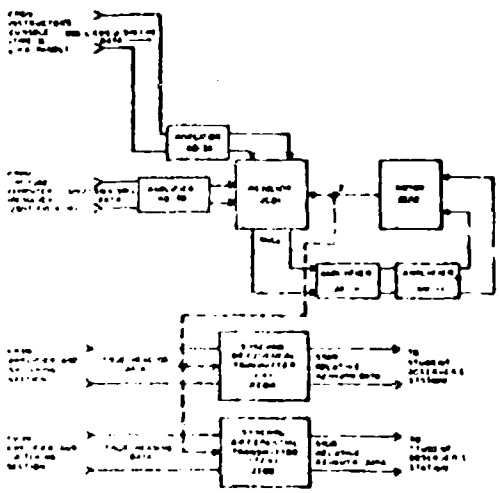


Figure 6 [illegible]

A SYNTHETIC TRAINER FOR TANK TURRET CREWMEN

By: Thomas Mongello, Project Engineer, U. S. Naval Training Device Center

Introduction - A need exists for a trainer for tank turret crewmen to provide complete synthetic training without resort to tactical equipment. A training device intended to meet this need has recently been developed for the U. S. Marine Corps by the U. S. Naval Training Device Center under Contract with the Jam Handy Organization, Detroit, Michigan.

The device is designed for use inside an enclosed building. It will permit realistic training to be carried out without interference from weather or dependence upon the availability of tactical tanks. It may be used on an around-the-clock schedule without requiring large tracts of land for firing ranges and with a minimum of the preparation necessary when training with operational equipment.

Description - The main components of the device, shown in Figure 1 are a simulated M47 tank turret mounted so that it rotates on a specially designed stand; a ten-foot radius hemispherical screen and a terrain projector which projects wide angle still photographic slides of real targets and back grounds; tracer, burst and point of aim projectors and an instructor's console which controls all the functions of the trainer and registers proficiency of the crew.

The device will serve to improve crew proficiency in the many skills and tasks required to expedite the training of tank turret crewmen.

The simulated turret is constructed of an open welded-pipe framework which permits crew activity to be viewed from outside the turret. Snap-on canvas covers are supplied to close the openings for more advanced training.

The turret interior is fitted with all the operational gear, controls, and sighting instruments used in the tactical tank. The main gun has been modified to recoil when the special ammunition developed and furnished with this device is fired. The special 90mm training round is designed so as to separate when fired. The projectile fells through a slot in the barrel while the cartridge case is thrown backward out of the breech.

The projector, mounted on a stand behind the turret, projects a still 60° picture on the inside of a hemispherical sc

The projected picture is photographed on a 3¼" x 4" lass slide plate. The photography for this trainer included tanks and other tactical targets. Twelve slides are mounted in a disc as shown in Figure 2. These slides may represent successive steps in the solution of a given tactical problem.

Up to five targets may be provided on each slide. The inclusion of more than one target per slide creates a problem of determining which target is of higher priority strategically. The ability to select the proper target is registered on a bank of lights on the console.

The twelve slides are carried on a disc mounted to the projector mechanism. The slide exposed is controlled by the problem number switch on the console. The trainer will accommodate a total of fifty problems or slides.

The slide also contains a projected reference spot for each target. This spot, not visible to the trainee with hatch closed provides the appropriate azimuth and elevation reference data for each target. The provision of this reference information as a projected spot rather than a coding on the slide permits errors to occur in the positioning of the slide during projection without affecting the accuracy of the scoring system.

The targets, terrain, and the burst and tracer displays are projected on the interior of a hemispherical screen 12 feet in radius. The turret rotates about the vertical center of the screen. The horizontal axis of the range finder passes through the center of the screen.

The operational stereoscopic range finder is modified to permit ranging on the projected targets which are actually at the same distance of 12 feet. This depart. from the principal of the range finder which is based upon binocular viewing points, in which one eye tends to see more of one side of an object than the other eye.

Three optical modifications have been made to the range finder to permit the simulated ranging to be carried out. These modifications are shown in Figure 3.

(1) The base length was reduced from 60 inches to six inches.
(2) An objective lens has been added to each telescope to change the focusing distance from infinity to 12 feet.
(3) A servo-driven optical compensator has been added to introduce synthetic range changes.

The objective lens added to achieve the 12 foot focusing distance also serves as the optical compensator and is servo-driven from the true range data in the console. The motion of the optical compensator is a function of the reciprocal of true range.

The modified range finder functions as follows: The acquisition of a target automatically selects the true range data from the stored data chassis within the instructor's console. The true range servo of the optical compensator lens drives to a position corresponding to the true range voltage. The compensator lens is then displaced an amount proportional to the actual range of the target then in telescopic view. The gunner must now rotate the range knob to correspond to the range of the target so that the target can be

made coincident with the range finder pips at infinity. The rotation of the range knob moves the range finder internal compensator lens and corrects the deviation back to infinity. The movement of the optical compensator lens takes place quickly, unnoticed by the gunner.

Additional modifications have been made to the range finder in order to feed data to the scoring system. This includes an ammunition number potentiometer pickup which feeds a repeater gyro mounted on the scoring console.

The burst, point of aim, and tracer projectors are mounted on a plate on the upper surface of the range finder, as shown in Figure 4.

The burst projector gives immediate indication to the trainee of his success in hitting the target. This indication is in the form of a momentary adjustably timed spot of light of constant size projected on the screen simulating a burst at a range of about 2000 yards.

The trace projector simulates the visual effect of tracer ammunition as a point in space where the real tracer burns out. The trace appears when the trigger is operated but lasts only momentarily. It is driven in elevation only and its position is a function of superelevation and ammunition number.

The burst projector is driven in both azimuth and elevation. The elevation position is a function of the difference between the estimated superelevation and the true superelevation, or the superelevation error. The azimuth position is a function of the cant angle. For simplicity, the effect of cant is simulated by tilting the projected picture in the opposite direction.

The instructor's console is shown in Figure 5. It contains all the controls by means of which the instructor presents the training problem and evaluates crew performance. The panel arrangement was based upon human engineering design considerations. The console also contains the target data storage system, the scoring detectors and computers.

The left side panel, shown in Figure 6, is the checkout panel. It contains a series of indicator lights representing before and after operation checks corresponding to the actual operations required to be performed in the tactical vehicle. This provides a means of evaluating the turret crew in the performance of these checks. The checkout panel also contains a timer, to check crew proficiency.

The center panel, shown in Figure 7 is the scoring panel. This panel contains the counters and lamps which enable the instructor to evaluate the crew performance in each problem presented. They indicate:

Problem Number
Rounds Over Range
On Range
Short Range
Rounds Left Azimuth
On Azimuth
Right Azimuth
Rounds Fired
Hit Scored

Each time the trigger is depressed, the rounds fired counter registers on the appropriate range and azimuth counters. Simultaneous on range and on azimuth registers a hit. The scores received evaluate the proficiency of the crew.

The scoring tolerance in ranging and aiming (azimuth and evelevation) is adjustable by means of switch settings on the face of the scoring panel. Credit is given for correct range based upon 1, 3, 7, and 15 units of error depending upon the range error setting used and for 0.5, 1.0, and 3.0 mils depending upon the aiming error setting. The aiming error scoring tolerance is always based upon a 3000 yard target. The unit of error as a measure of error in range is based upon one unit of error being equivalent to 4.65 times the range in thousands of yards squared or

$$\text{Unit of Error} = 4.65 \left(\frac{R}{3000}\right)^2.$$

Information for determining the aiming error is obtained from a Kollsman target tracker.* It contains a telescope and photoelectric scanning mechanism positioned in azimuth and elevation by restoring motors and is capable of locking on a projected target spot. The tracker is mounted on top of the range finder and moves with the turret, fixing on a spot when it "sees" one. The target tracker is designed with a fixed elevation offset to correspond with the exact angle of offset of the spot printed on the slide for each target. Azimuth offset angle is zero. When aiming is correct both in azimuth and elevation, the error signals generated by the target tracker are zero. Any error in aiming in either azimuth or elevation produces an error signal which is used to determine the accuracy of scoring.

The scoring panel also contains an ammunition number dial which indicates the ammunition number set on the control knob in the range finder.

A target selected sequence indicator also provided on the scoring panel and shown in Figure 7 gives additional information on crew performance. It consists of a bank of 25 lights which indicate when the trainee is aiming within the 2½ degree cone angle of the target tracker (by the proper indicator lighting dimly); the priority of the target #(A,B,C,D, or E); when the 90mm gun has been fired at this target (by the

*Kollsman Type No. XA-1872-9980

same indicator lighting brightly); rounds fired at each target; and the order of shots fired (1 through 5).

The panel on the right, Figure 8 [illegible] power panel. It controls trainer power, [illegible] power, projector display, simulation effects [illegible] communications. The simulation effects are hangfire, recoil, 90mm gun sound, machine gun sound, and engine noise.

The 90mm gun sound is on a closed loop of ¼" magnetic tape and is set to cycle for one round. Machine gun and engine noise are on the same tape. The machine gun noise is on a seperate track and is played only when the trigger is operated. All sound effects are played on a speaker located in the turret.

The power rack shown in Figure 1 receives single-phase 60 cycle power and provides all power with the exception of turret power and terrain projector lamp power. This rack supplies 300 volts and 24 volts D.C.; 6.3 volts and 110 volts, 60 cycle and 115 volts and 26 volts 400 cycle to the instructor's console.

Theory - The trainer operates as an analog computer in which ranging and tracking accuracies or aiming errors are represented as voltages, and the direction of error is represented by the phase angle of the voltage. One mil is represented by a value of 0.46 volts. All the servo systems and phase detectors operate on a 60-cycle carrier system.

The azimuth, elevation, range and cant data for fifty problems is contained on 800 carbon potentiometers located within the stored data chassis in the instructor's console, shown in Figure 9. When the instructor switches to a particular problem number pertaining to a slide, he automatically by operating a multiple 50 point stepping switch selects a group of 16 of the 800 data potentiometers pertaining to the slide selected. This provides azimuth, elevation and range data for each of five targets, and the cant for the entire problem. The next selection is done automatically by the azimuth and elevation data detectors shown in Figure 10.

These detectors compare azimuth and elevation position signals which are received from the turret with the stored course azimuth and elevation data of the five targets on the projected slide. The azimuth indicator has been modified by the insertion of two gears and two potentiometers to pick up turret position. A similar change has been made to pick up elevation elevation from the range finder. When the turret and gun position signals are close to the stored data signals of any target on the slide, the stored true range data for that target is fed to the scoring system and other components.

All range data voltages vary as a reciprocal of range to permit comparison with the range finder output which is also a reciprocal of range. The inputs to the computer are:

1. Turret azimuth
2. Gun or range finder elevation (taken from the range compensator servo section of the range finder).
3. Range finder estimated range as a reciprocal function.
4. Superelevation angle as computed by the range finder.
5. The ballistic-ammunition setting of the range finder - (taken from a potentiometer located in the range finder which takes the sum of the ammunition selector knob position and the ballistic correction knob position.)

The range finder superelevation angle is fed through summing resistors to the burst elevation and tracer elevation servos. At the input to the tracer elevation servo, the range finder superelevation angle is summed with the ballistic correction and ammunition number signal and with the positive feedback signal from the tracer elevation servo drive. The range finder superelevation angle signal is also summed with the true superelevation angle signal and the position feedback signal from the burst elevation.

The true range signal expressed as a reciprocal is fed to four units of the trainer as follows:

1. The range finder optical compensator true range servo.
2. The true superelevation computor true range servo.
3. The over range phase detector (Shown in Figure 11).
4. The under range phase detector (Shown in Figure 11).

The true range signal is repeated by the feedback potentiometer on the superelevation true range servo and is fed to the aiming tolerance selector switches in two forms (Phase A and Phase B). This makes the aiming tolerance voltage a reciprocal function of range with large aiming tolerances at small ranges and small aiming tolerances at large ranges.

Azimuth and elevation sighting error is obtained from the target tracker synchros which generate voltages proportional to the sighting error and of the proper phase. The sighting error voltages are fed through two azimuth and two elevation phase detectors and compared with the voltages representing the tolerances set in on the console.

The true superelevation computer is shown in Figure 12. It has two inputs, reciprocal of true range and ballistic correction ammunition number, the same as the trainer range finder.

The computer operates on the same principle as the trainer range finder using a ballistic cam identical with that of the range finder. It computes true superelevation which is fed to the burst elevation servo system. It also computes the product of superelevation angle and cant angle. This product is fed to the burst azimuth servo through a summing resistor. The burst time and tracer time are also computed here.

Many of the more conventional circuitry functions such as checkout procedure, operation of the phase detectors, details of the data storage system, operation of the target selected sequence indicator are not described due to limited space.

Conclusions - The novel features of this trainer are:

1. It is the first synthetic device for training tank turret crewmen in all phases of turret operation with ability to check and evaluate performance.

2. The use of a target tracker to establish an accurate reference line from reference data on a film.

3. The modification of a stereoscopic range finder to range on photographic targets projected on a screen.

Credit - Appreciation is extended to Mr. John Cambell and Mr. Dana Manning and others of the Jam Handy Organization for their work in the design and development of this unique trainer.

Additional Information - Additional information on this trainer is given in U. S. Naval Training Device Center Maintenance Handbook for Device 3T1, NAVTRADEVCEN-P-1884.

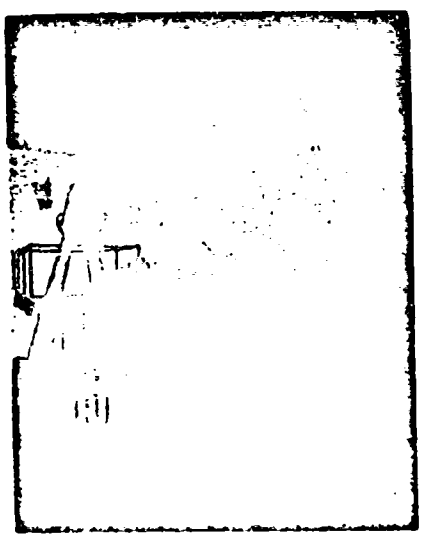


FIGURE 1
Trainer Assembly
1. Target Projector
2. Turret
3. Screen
4. Power Rack
5. Console

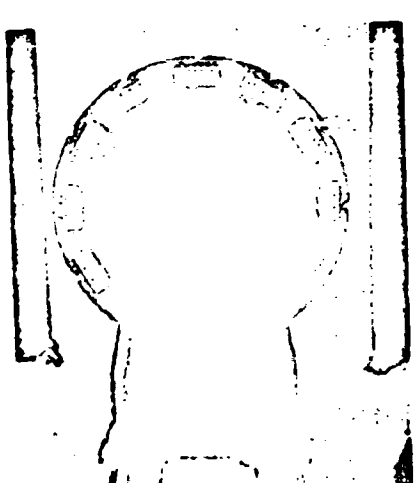


FIGURE 2
Terrain Projector

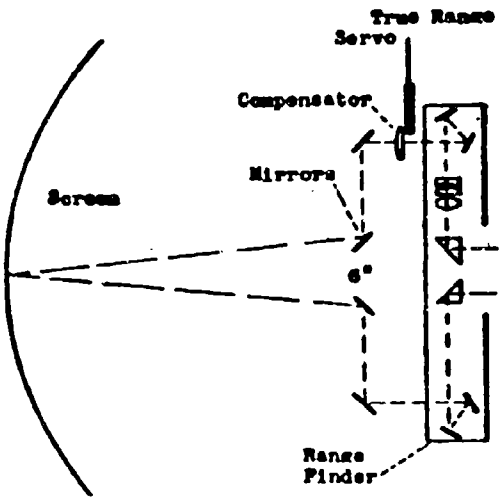


FIGURE 3
Range Finder Modification

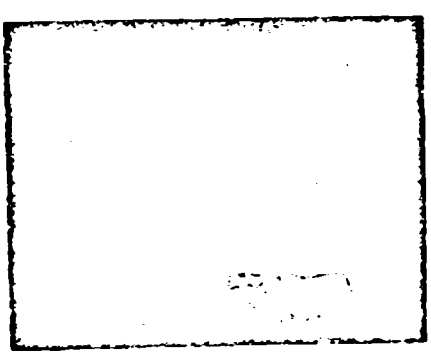


FIGURE 4
Range Finder and Fire Simulation

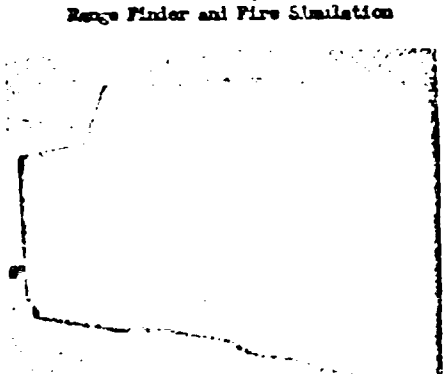


FIGURE 5
Instructor's Console

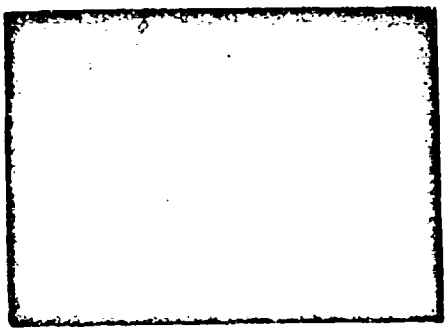


FIGURE 6
Checkout Panel

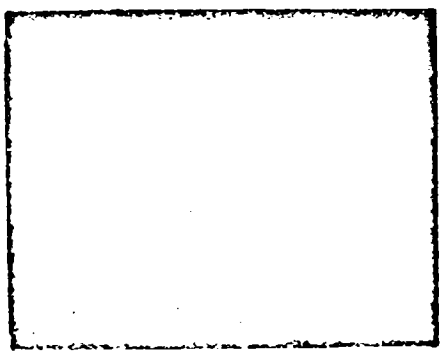


FIGURE 7
Scoring Panel

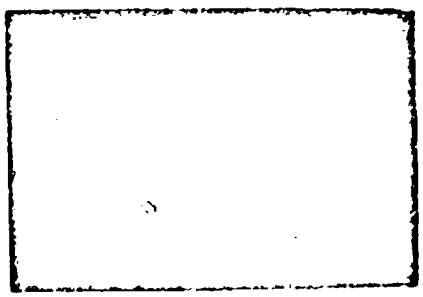


FIGURE 8
Power Panel

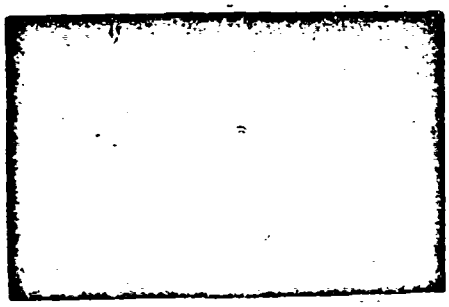


FIGURE 9
Stored Data Chassis

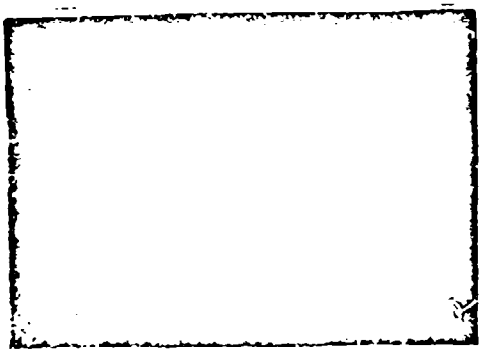


FIGURE 10
Data Detector Chassis

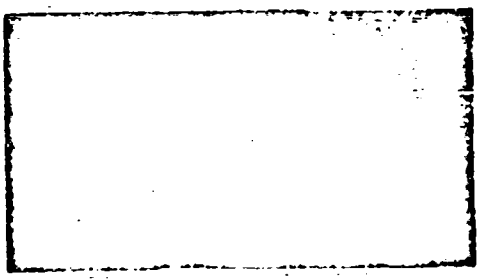


FIGURE 11
Phase Detector Chassis

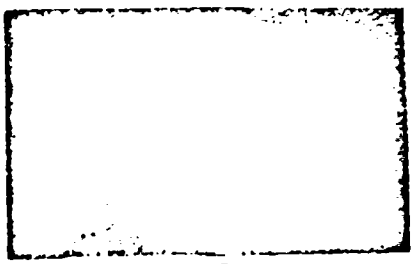


FIGURE 12
True Superelevation Computer

MAINTENANCE, REPAIR, AND ASSEMBLY IN SPACE BY REMOTE MEANS

By: Dr. John W. Clark, Nuclear Electronics Laboratory, Hughes Aircraft Company

Introduction

Our determination to explore space, the moon, and (later) the planets, and to develop scientific knowledge which may lead to commercial and military operations almost unimaginable today is a challenge requiring most creative application of science and technology. This paper is concerned with remote handling, which is one of the many technical disciplines which collectively make up "Space Technology".

Modern remote handling technology is being created by a combination of electronic control engineering as developed for guided missiles and radar, of human engineering as developed (mainly) for aircraft cockpit design, of remote handling as the nuclear industry is beginning to employ it, plus numerous others.

Machines and systems can now be built which enable a man to work in an area which he cannot enter. Such systems have numerous applications in space.

Remote Handling Vocabulary

The new technology of remote handling and operation in hazardous areas is sufficiently unfamiliar as to justify a preliminary discussion and a few basic definitions before exploring in more detail the applications of this technology to space programs.

In most general terms consider Figure 1, which shows a hazardous area in which one desires to perform some operation. This area may be hazardous due to lack of air, the presence of nuclear radiation, to high temperature, or to any other circumstance which makes it impossible for a man to enter. The operations required may be equally broadly considered. Typically, it is necessary to translate objects from one position to another, to operate tools, such as screw drivers or wrenches, or to operate measuring equipment of many kinds. Usually, fixed obstacles are contained within the hazardous area, limiting the freedom of motion of any equipment contained therein.

A typical problem is that of removing a spark plug from a modern automobile engine. Visualize the requirements upon a mechanism which can do this if the entire automobile is contained within a hazardous area which cannot be entered by personnel.

In the past, problems of this type have usually been attacked by the use of long tongs or forceps. Some of these have become extremely complex and have employed hydraulic or electromechanical actuators to supplement the physical strength of the operator. All such devices may be considered extensions of the man's arms, and the man himself is visualized as just outside the hazardous area and possibly protected from it by a suitable barrier or shield.

Let us approach this problem in a more generalized way, and let us assume that the separation between the hazardous area and the safe area is so great that tong-like tools cannot be used. This requires us squarely to face the problem of operating in areas which are inaccessible. This is sometimes done by providing the operator with protective clothing suitable to the environment. Deep-sea diving suits serve this purpose in the underwater environment. Movable personnel shields may be used in nuclear environments, and the space suit of science-fiction is often proposed for the space environment.

In all such cases we find that we have actually not solved the problem. We have, on the contrary, merely changed the geometry. The safe area is now contained within the hazardous area, but the operator is still separated from his work by his protective system and must use long tongs rather than his own hands for manipulation.

From the viewpoint of solving the fundamental problem of operating within an inaccessible as well as hazardous area, let us analyze a man as a handling system. For this purpose we may ignore many of his more interesting attributes and note that just four interrelated systems are involved. These may be identified briefly as his brain, his eyes, his hands, and his feet. More seriously, the eyes and the sensory nervous system provide information concerning his surroundings. The hands, as controlled by the motor nervous system, are able physically to move objects as desired. The feet and legs, controlled by their separate motor nervous system, enable the entire organism to move about; and finally, the brain assembles and organizes all these data and directs the motor systems as required.

The three functions symbolized by eyes, hands, and feet may readily be extended by modern electronic means over any desired distance. In this way, a man's senses and his ability to accomplish useful work may be extended to any distance, while his brain, which can be duplicated by no existing computer, remains in safety and comfort at any desired location.

This simple concept is the basis of the new technology of remote handling. Remote handling systems ("Mobots") are not to be looked upon as competitive with Man-in-Space; rather, the effective utilization of such systems will increase the ability of the man to accomplish functions in space. This paper is concerned primarily with a preliminary outline of ways in which the concept of remote handling systems can be applied to several realistic space situations.

Analysis of Generalized Remote Handling System

A block diagram of a generalized remote handling system is given in Figure 2. It is desirable to consider this block diagram prior to consideration of specific space systems in order to clarify the several fundamental systems of which

these systems must be built, and in order to clarify the engineering analyses required for design of such systems.

As noted above, remote handling systems are artificial extensions of man's senses and muscles over considerable distances. In order to accomplish this in addition to the systems which duplicate the senses and the muscles, a link must be provided actually to bridge the physical gap between the man and the remote machine, and a control console must be provided with which the man communicates with the remote machine.

All of these are shown in Figure 2 as individual blocks in the system diagram.

As a further addition to the vocabulary of this topic, the term "Mobot" has been coined to refer to the mobile remote portion of the system. Mobots are mechanical units and are the most conspicuous portion of remote handling systems in action. One basic psychological observation can be made, namely, the ease with which a Mobot operator learns to identify himself with the Mobot and to forget the existence of the interconnecting systems. This psychological identification appears to be a basic necessity of successful operation of fully-remote handling systems.

Let us consider briefly the requirements upon the several basic subsystems.

Hands and Arms. It seems impossible to avoid the anthropomorphic terms, hands and arms, to refer to the manipulating devices. In general, however, these do not much resemble human hands or arms, but are designed specifically to perform tasks as required. It is extremely difficult to rival the versatility of the human hand. However, special-purpose handlers can usually out-perform the human hand in either dexterity, strength, small size, or other specific attributes. In addition to having sufficient strength to handle the assigned tasks, Mobot manipulating systems must be able to work in the presence of obstacles and to perform complex and intricate motions. Remember the automobile spark plug as a typical operation which one might wish to remotize.

The Senses. While all the human senses can rather readily be transmitted via electronic means, vision is the most important by far, and the only one which will be discussed in this brief analysis.

Spatial orientation is normally accomplished by a variety of methods. Parallax, scale, relative motion, and the like, are probably most important of these. Binocular vision is surprisingly unimportant, as demonstrated by the fact that one-eyed men are but little handicapped in perceiving spatial orientations in their vicinity. Based on this analysis, excellent success has been obtained with a simple vision system utilizing two TV cameras, as shown in Figure 3. These two cameras show the operator two mutually perpendicular projections of the area viewed, from which he can learn to deduce the spatial orientation of all objects within his visual field. Learning time of a few hours has proven quite adequate for this system.

Locomotion. In order for the Mobot to move freely about in space, locomotion is best provided by auxiliary jets suitably located to provide translation and rotation. On lunar or planetary surfaces locomotion may be provided by conventional wheels or by quite unconventional walking devices which become practical with the use of Mobot control circuitry. For operation in and around large satellites, locomotion is readily obtained by grasping hand-holds or other protuberances with the same arms utilized for handling operations.

Command and Data Link. This link is the communication between the operator and the Mobot. It must obviously be capable of providing command information at a sufficiently rapid rate to control the Mobot, and of transmitting sensory data from Mobot to operator, again at an adequate rate. For short distances a cable may be used, which can also serve to transmit operating power to the Mobot. For longer distances conventional radio links are suitable. Considerations of data rate and trade-off between bandwidth, antenna dimensions, and power required are the principal system requirements.

Power Requirements. Prime power for the Mobot in systems utilizing cables is usually brought in via cable. In completely remote systems utilizing radio command and data link, prime power may be furnished by any of the familiar power sources. The selection of an appropriate source is of course another basic step in the over-all system design. Radioactive power sources, such as the "SNAP" series, are particularly appropriate for small space systems; a reactor may be required to power larger space Mobots.

Control Console. This is the man-machine link. Principal requirements in its design are operator comfort and convenience. After a considerable learning period, an experienced operator becomes completely unaware of the console and subjectively identifies himself with the Mobot. The console must be designed in such manner as to facilitate this identification.

Mobot Experience

In order to demonstrate that the concept of a fully-remote mobile remote handling system is indeed practical, let us refer briefly to experience with Mobots for the nuclear environment. One such machine is Mobot Mark I, shown in Figure 4. This machine was built for use in a nuclear radiation laboratory and has now demonstrated many months of successful operation. Such operations as pouring liquids, opening and closing doors, operating small power tools, and the like, were readily accomplished with this machine. Learning time varies from 4-8 hours, and real facility is attained after about 40 hours.

Another nuclear Mobot is Mark I-B, shown in Figure 5. This Mobot is designed primarily for use in nuclear hot cells. It is a modular design and can be assembled in many different configurations to meet different requirements. The reason

for mentioning it here is to show another example of a fully-remote handling system which has been successfully operated for some time.

Applications of Remote Handling Systems in Space

It should be clear from the preceding discussion that there are numerous applications in space programs for fully-remote handling systems. These are tabulated in Figure 6. In considering these systems, such matters as timing, payload availability, and the like, are relevant, as well as the over-all technical possibilities. The lunar orbit is a convenient dividing line for these considerations. The balance of this paper will give separate consideration to the applications of these systems to cis-lunar, lunar, and trans-lunar applications.

Cis-Lunar Applications. In the space between the exosphere and the lunar orbit, remote handling systems can be applied in the immediate future to maintenance and repair of small satellites; and in the more distant future, for assembly, maintenance, and repair of large satellites.

Figure 7 shows an artist's concept of a Mobot performing some routine maintenance function on a small satellite. A manned orbiting vehicle, such as the proposed NAC-LAC Space Ferry, is being utilized as the control point for this function.

The satellite itself may be a communication satellite, a weather satellite, a scientific satellite, or any satellite performing function of this type, the cost of which is sufficiently great as to justify maintenance by such means.

The man in the space ferry controls the Mobot through a cable of moderate length, perhaps a few hundred feet. Thus, the problems of command and data link and of prime power are relatively simple, and the Mobot can perform rather complex operations, such as replacing vacuum tubes, readjusting circuits, or the like, as may be required. Upon completion of its operations, the Mobot detaches itself from the satellite, is returned to the space ferry, and the entire system returns to earth. This completely eliminates the need for a space suit or separable capsule, while at the same time increasing the value of the space ferry pilot's presence in space.

Turning to the more distant future when it becomes necessary and desirable to build orbiting platforms too large for a single booster, a different type of Mobot then becomes extremely useful. As shown in Figure 8, this Mobot can retrieve large objects which have been orbited separately, can bring them together, and actually join them with suitable bolts or fasteners. This operation can be controlled from a manned vehicle as in the previous discussion, or, alternatively, it can be controlled from a ground station utilizing communication satellites to relay the command and data link information to all points in the Mobot orbit. After these large space stations are assembled, Mobots are still useful in any and all operations on the outside of the space station, at which time they would presumably best be controlled by men in the space station rather than by men on the ground.

Lunar Applications. In considering remote handling systems for lunar application, the time schedule becomes all important. The development of remote handling systems should logically parallel that of booster rockets, so that as the thrust and payload available increases, remote handling systems to take advantage of it are available.

In the immediate future it will be possible to deliver to the surface of the moon payloads in the vicinity of a few hundred pounds. The scientific data obtainable with such payloads will be greatly increased if extremely simple remote handling systems are employed to set up the instruments and to place samples in them. One visualizes for this purpose something like the device shown in Figure 9. This contains a single arm with very limited motion, intended to gather a sample of the lunar surface and to obtain simple data concerning its mechanical properties. This rudimentary Mobot is controlled from the earth by a radio link. Preliminary calculations verify this to be quite feasible. As the payload available increases, the functions accomplished by the Mobot can correspondingly increase until quite sophisticated physical, metallurgical, chemical, and mineralogical instruments can be employed and suitable samples gathered for each.

Still further in the future are complete free-moving Mobots which can move about the lunar surface independent of the delivery vehicle. Since we do not yet know the exact nature of the lunar surface, it is difficult to predict the appearance of these vehicles at this time. One possibility is shown in Figure 10 of an unconventional method of locomotion suitable for extremely rough terrain.

The availability of vehicles of this type implies the possibility of delivery of material to the lunar surface in a multiplicity of separate rockets, and utilizing Mobots to assemble complex lunar installations. In other words, a lunar base may be possible prior to a manned lunar landing. In this connection, it is noteworthy that the total payload required for a quite complex Mobot may be considerably less than required for a man, particularly in view of the fact that a Mobot will make no objection to a one-way trip.

This indicates the interesting possibility of preparing a lunar base prior to the first manned lunar voyage, so that the man or men who make this trip will find some base facilities to aid their landing and, most important, their return to earth.

By the same token, the availability of Mobots will greatly increase the effectiveness of the men who make the early lunar explorations, since it will eliminate the necessity for these men to leave their safe and comfortable rocket capsule in order to explore the lunar surface. Mobots which may have been transmitted earlier by

separate rockets will enable them to gain all the information desired without the necessity of leaving their vehicle.

Trans-Lunar Applications. The application of the Mobot concept to planetary exploration is somewhat different due to the enormous distances involved. One feels no great reluctance in contemplating a lunar command and data link; however, a planetary one does not appear feasible even on a 20-25 year time scale. However, Mobots will undoubtedly find their uses in planetary exploration at such time as manned expeditions to planetary distances become feasible. For example, it may well be sensible to plan a trip to Mars or Venus in which the space ship orbits the planet for a period, during which time a Mobot is delivered to the planetary surface to explore and obtain scientific data under control of the man in the space ship. Upon completion of the mission, the man would return to earth without having landed, thus conserving the considerable amount of fuel required to enter and leave the planetary surface.

The more distant planets will place more severe demands upon systems of this type, since their atmospheres are poisonous and the temperatures extremely cold. It is reasonable to suppose that by the time it is possible to reach the vicinity of Saturn and Jupiter, Mobots to explore their inhospitable surfaces will be quite practical.

A Possible Time Table

The several systems described above have an unavoidable Science-Fiction flavor. Nevertheless, all are perfectly practical in the sense that known and proven engineering principles are used. These are programs which can be accomplished by diligent application of intense effort; it is not necessary to postulate either inventions or scientific breakthroughs to ensure success.

Perhaps most significant is the demonstration that it is indeed possible to perform complex operations with a mechanism electronically linked to a man; Mobot Mark I clearly proves this point.

The time required to develop and construct space Mobots is comparable to that required to develop high-thrust boosters and the other major components of a space vehicle. Thus, simple Mobots can be ready in time for instrumented soft lunar landings; and versatile, freely moving Mobots can be completed about the same time that the necessary payload of a few thousand pounds is available.

In evaluating the significance of remote handling systems in exploring and developing the moon, it is important to remember that a Mobot will accept a one-way ticket to the moon (which a man would probably be reluctant to do); and it can be designed to operate in vacuum and over a considerable range of temperature. These facts permit one to deliver quite a versatile Mobot to the moon even before payloads are adequate to deliver a man there and return him to earth. One recognizes the political and psychological value of actually placing a man on the moon; remote handling systems and Mobots can expedite his arrival there, and improve his effectiveness when he arrives.

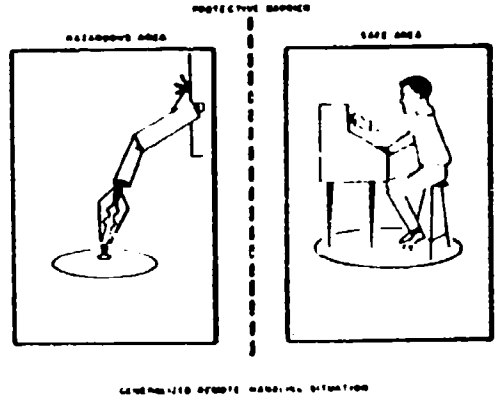


FIGURE 1. GENERALIZED REMOTE HANDLING STATION

Hazardous Area

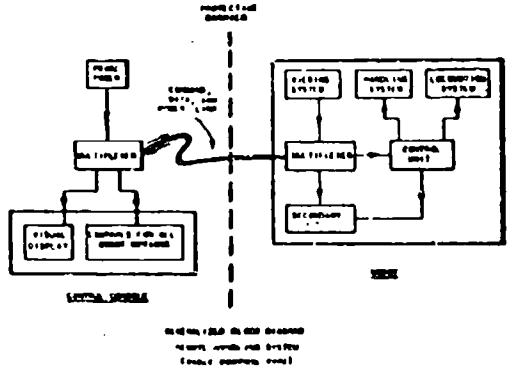


FIGURE 2. GENERALIZED BLOCK DIAGRAM

· Remote Handling System (Cable Control Type)

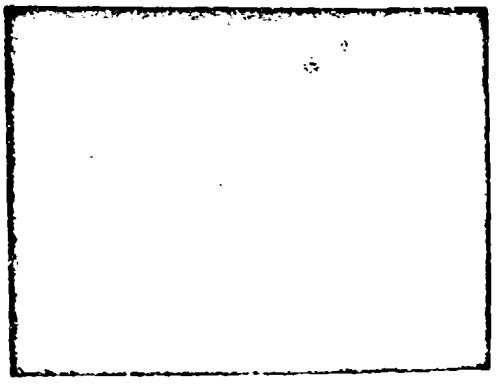


FIGURE 3. DUAL CAMERA TELEVISION SYSTEM



FIGURE 4. MOBOT Mark I - Control Console at Right

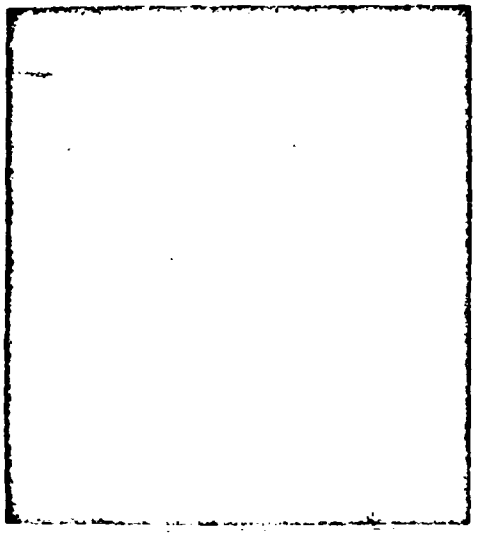


FIGURE 5. MOBOT Mark I-B

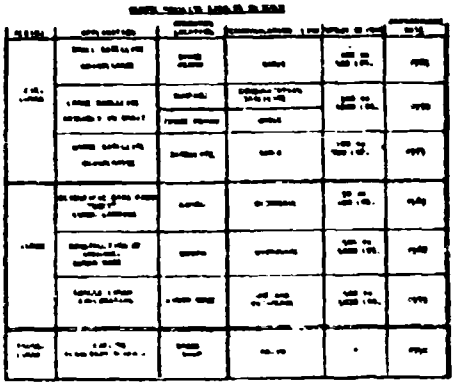


FIGURE 6. REMOTE HANDLING SYSTEMS IN SPACE Cis-lunar, Lunar and Trans-lunar.

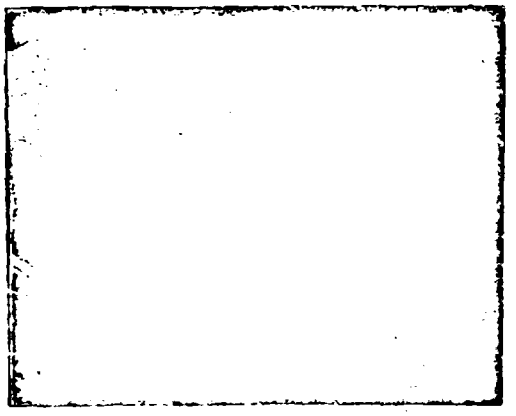


FIGURE 7. SPACE MOBOT REMOVING HATCH OF SATELLITE Space Ferry Control in Background.

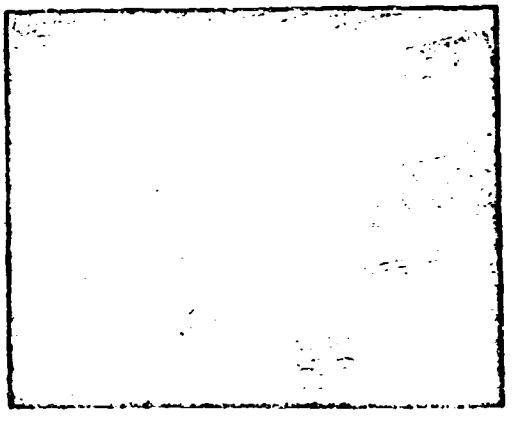


FIGURE 8. SPACE MOBOT ASSEMBLING PLATFORM

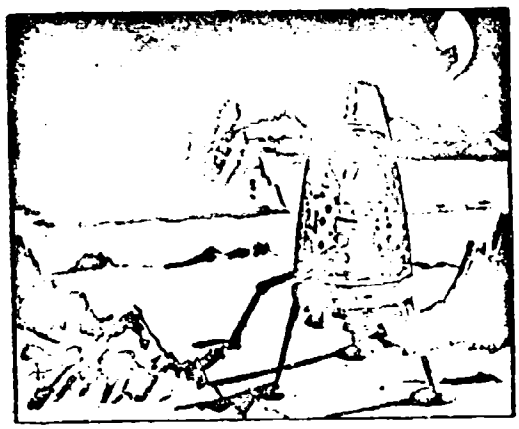


FIGURE 9. LUNAR ROBOT — Single Arm

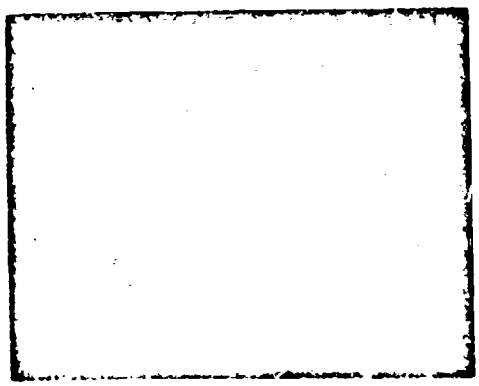


FIGURE 10. PENTAPEDS ON LUNAR SURFACE — Equipped with Robot Arms and T-V Cameras.

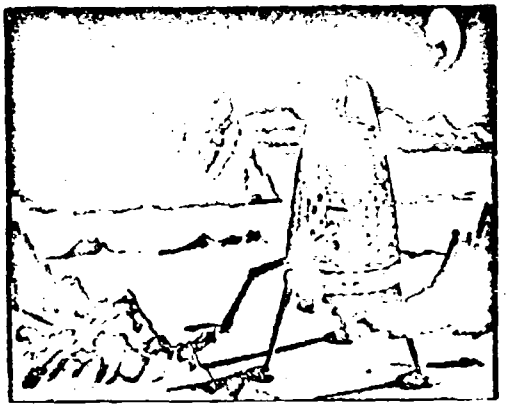


FIGURE 9. LUNAR ROBOT — Single Arm

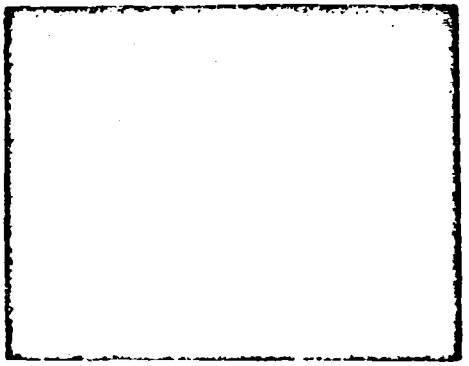


FIGURE 10. PENTAPEDS ON LUNAR SURFACE — Equipped with Robot Arms and T-V Cameras.

A FUNCTIONAL DESCRIPTION OF PIONEER V

By: Mr. Paul F. Glaser, Experimental Space Projects, Space Technology Laboratories, Inc.

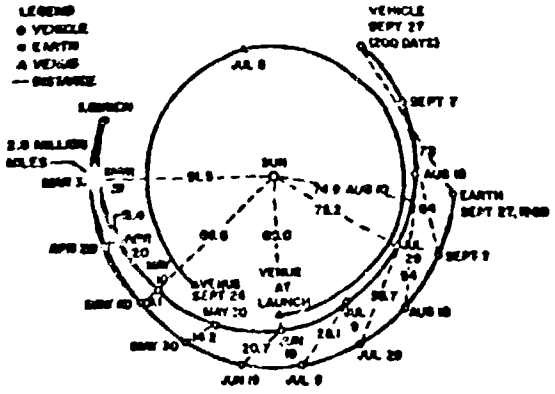
1. INTRODUCTION AND SUMMARY

The mission of the Pioneer V space vehicle is to explore space out to the maximum range of present communication facilities, in the range of 50 to 60 million miles. The payload, illustrated in Figure 1, was launched on March 11 1960



Figure 1.

into a planet-like orbit about the sun and on May 5, 1960 is about 7.6 million miles from the earth. The orbit of Pioneer V will carry it to perihelion on August 10, 1960, at a distance of 74.9 million miles from the sun. At that time the range from the earth to the payload will be 47 million miles. This distance will continue to increase until the maximum range at which communication is possible is exceeded, probably in early September (see Figure 2).



TRAJECTORY OF PIONEER V

Figure 2.

Pioneer V is the lightest of the three payloads in the recent Able series, 94.8 pounds compared to the 143 pounds of Explorer VI and the 389 pounds of Able-4. Pioneer V, however, carries no propulsion equipment and therefore a greater proportion of its weight can be devoted to instrumentation. This instrumentation consists of sensing and measuring equipment for the scientific experiments, the Telebit unit, a doppler transponder consisting of a command receiver and transmitter, a solar cell power conversion system including storage batteries and converter circuitry, and associated logic. A functional block diagram of the payload is given in Figure 3.

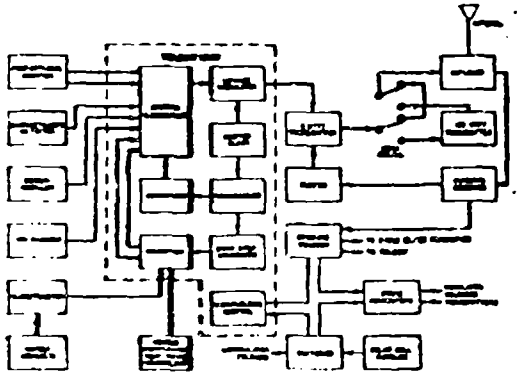


Figure 3.

The launch vehicle was nearly identical to that used for the Able-3, Explorer VI satellite. First stage was a standard Thor IRBM, with a modified autopilot control system. Second stage was an Aerojet AJ 10-1A with STL autopilot and radio guidance equipment. After second stage burnout, six spin rockets on the stage fired and spun up the third and fourth stages to approximately 2.4 rps. Third stage was an Allegany Ballistics Laboratory 248A4 solid propellant rocket. When this stage burned out, at an altitude of approximately 221 nautical miles, the vehicle velocity was 36,400 ft/sec. Since escape velocity at this altitude is about 35,555 ft/sec, both the payload and the third stage left the earth to enter elliptical orbits about the sun, although when payload and third stage were separated 20 minutes later, the slight velocity increment imparted by a separation spring has caused the two stages to drift apart.

2. STRUCTURE

Pioneer V is an approximate spheroid, 26 inches in diameter at the equator. Four radial extruded aluminum beams, forming the basic support structure, originate at an apex at the bottom of the payload and terminate at the

periphery of the central platform, which is constructed of fiberglass honeycomb. A top cover of formed sheet metal surrounds the upper section of the payload, and formed sheet metal panels cover the lower portion. The sheet metal in both cases is an aluminum alloy. All instrumentation attaches to the central platform, clustered so far as possible at the periphery to increase the roll moment of inertia. Four solar cell paddles, attached equidistantly about the equator of the payload at the points where the beams join the platform, fold down about the third-fourth interstage during launch, and were released and locked in their extended position immediately after second stage shutdown.

The 94.8-pound weight of the payload breaks down as follows:

Structure and covers	18.17
Wiring, potting, and connectors	3.86
Instrumentation	33.51
Experiments	9.36
Batteries	17.20
Paddles and solar cells	10.26
Temperature control material	0.70
Dynamic balance weights	1.74
	94.80

3. EXPERIMENTS

The scientific apparatus in the payload is designed to measure three types of phenomena in space throughout the life of the equipment: radiation, magnetic fields, and micrometeorite density. Location of the equipment in the payload is shown in Figure 4.

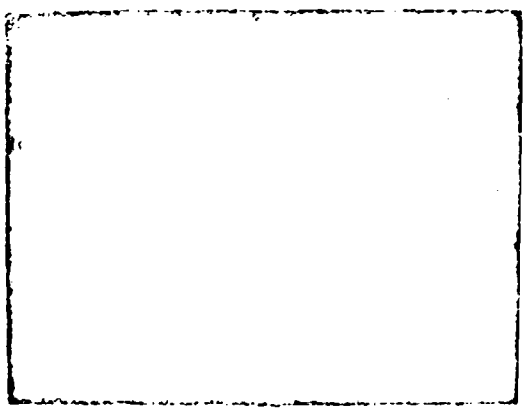


Figure 4a.

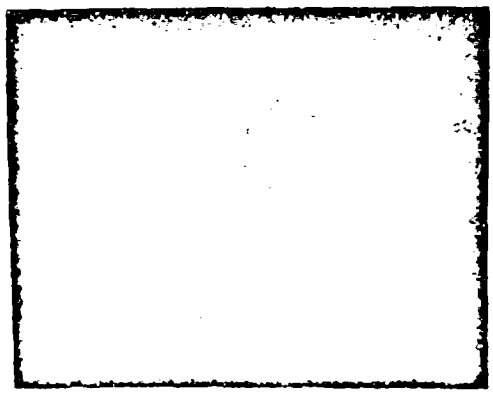


Figure 4b.

3.1 Radiation Experiments

Radiation is measured by two sets of experiments: a University of Minnesota ion chamber and Geiger-Mueller tube and the University of Chicago proportional counter telescope. This equipment is designed to determine the relative abundance of the different species of charged particles and to indicate their energy distribution.

The ionization counter in the University of Minnesota experiment is of the integrating type. It consists of a four-inch sphere filled with argon gas to an absolute pressure of approximately six atmospheres. As a central quartz rod collects electrons formed by ionizing radiation in the argon gas of the chamber, its potential drops to the point where a charging pulse is actuated. Each pulse from the chamber represents the collection of approximately 10^{-11} coulomb. The time interval is determined by allowing two ion chamber pulses to gate a clock generator; the clock pulses are stored in the telemetry unit and read out when the transmitter is operated.

The second detector used by the University of Minnesota is a small Halogen Geiger Counter (Anton Type 302). The data is handled in the same manner as the ion chamber except that a scale factor of 2^{10} is used and the time interval to reach a 2^{10} count is measured.

The proportional counter telescope provided by the University of Chicago contains a bundle of seven small proportional counters arranged as a central counter surrounded by two sets of three counters. The counters are filled with a mixture of argon and methane to a pressure of about 30 centimeters of mercury. The central counter is connected to an amplifier with a threshold of about 1 millivolt; the amplifier operates a scaler chain of 2^{11}. The output of each of the outside groups is applied to a separate amplifier, also with 1-millivolt threshold. The output of all three amplifiers is used to

activate a triple-coincidence circuit which in turn is applied to a 2^9 scaler chain.

As a result of this geometry, the counter sensitivity is almost isotropic. A singles event corresponds to a charged particle traversing one outer counter and entering the center counter. Such an event is recorded only on the 2^{11} scaler chain. A triples event corresponds to a charged particle traversing one outer counter, the center counter, and either into or through a counter in the other group of three. Such an event is registered on both the 2^{11} and 2^9 scaler chains. The minimum energy required to penetrate the triple coincidence counter is 70 mev for protons and 12 mev for electrons.

3.2 Magnetic Field Measurements

The search-coil magnetometer carried in Pioneer V measures the component of the magnetic field perpendicular to the spin axis. By making use of the spin of the vehicle and knowledge of the orientation of the axis of this spin, the magnitude of the magnetic field through which the satellite is moving can be deduced. A phase comparator which measures the phase relationship between the output of a photodiode sun scanner, or aspect indicator, and the output of the magnetometer permits accurate knowledge of the direction of the field. The sensing element of the magnetometer consists of a mumetal core wound with 5000 turns of No. 4 copper wire. The output of this coil is coupled to a transistorized amplifier tuned to the nominal spin rate of the vehicle.

3.3 Micrometeorite Detector

The micrometeorite momentum spectrometer for the Pioneer V system determines the times of impact of micrometeorites and separates the momenta into two groups of different energies. Equipment consists of a diaphragm on the payload shell, a microphone mounted beneath the diaphragm, an amplifier, and a pulse height analyzer. A particle striking the diaphragm transfers its momentum inelastically to the diaphragm, and the 100-kc component of the impulse is picked up by the microphone by means of a piezoelectric crystal which resonates at this frequency. A two-channel pulse height analyzer classifies the amplified pulse according to whether it was produced by a low momentum (less than 3×10^{-3} gm-cm/sec) particle or a high momentum (greater than that value) particle. Minimum sensitivity of the detector is about 10^{-4} gm-cm/sec.

3.4 Astronomical Unit

In addition to the scientific results made possible by the equipment carried specifically for that purpose, the precise long-range tracking data will permit an accurate determination of the length of the astronomical unit. This unit, the semi-major axis of the earth's orbit, is a basic astronomical constant. It is the unit for determining the absolute distances of solar system bodies from each other; it is the basis for the parsec unit of astronomical distances; its magnitude, and the derived value of the mass of the sun, are fundamental to all space trajectory calculations. It is felt that the AU is now known to within about 0.04 per cent. Tracking of Pioneer V over long distances is expected to improve this value. Depending upon the length of time the vehicle is tracked, a determination of the AU to within 0.03 per cent or better is anticipated.

4. INSTRUMENTATION

4.1 Communications

Transmitting and receiving equipment within the payload permits two-way communication. The payload transmitters (a 5-watt and a 150-watt unit) convey telemetry information to the earth and a high-sensitivity receiver permits the reception of earth-transmitted commands. When interconnected coherently, the payload receiver and transmitter form a transponder capable of providing both range and range rate information for tracking the payload.

For communication to the earth, the transmitter accepts an RF signal from the payload receiver, multiplies it eight times in frequency, and amplifies it either to a 5- or 150-watt level. (When the 150-watt transmitter is operating, the 5-watt transmitter acts as a driver.) In the process the signal, whose frequency is 378 mc, is phase modulated with a 1024-cps subcarrier containing the time-multiplexed pulse-code-modulated output of the Telebit system. Biphase modulation is employed to impress the telemetry output on the subcarrier. The transmitter weighs about six pounds, and has an efficiency of about 15 per cent when operating at 5 watts and about 30 per cent when operating at 150 watts.

Two means are used in Pioneer V to provide for maximum telemetry and tracking information during its lifetime within its weight, power generating and storage system, and telemetry system capabilities. The two transmitters allow use of the low power 5-watt unit while it is close enough to the earth and thus maximize the transmitting time within the power system capabilities. The high power 150-watt transmitter can be used in lieu of the 5-watt transmitter when range becomes sufficiently large to prevent ground stations from receiving data on the low power unit. Transmission time is reduced accordingly, although up to the point where the 150-watt unit is required maximum transmitting periods are assured.

As range increases bandwidth constraints become more evident with either the 5- or 150-watt transmitters. By using a digital telemetry system it is possible to narrow the information bandwidth by dropping the information bit rate from 64 pulses per second to 8 pps and finally to 1 pps with increasing range.

The variations in transmitter output power and information bandwidth are accomplished by ground command and all combinations will be used during the active life of Pioneer V to insure that maximum channel capacity and transmission

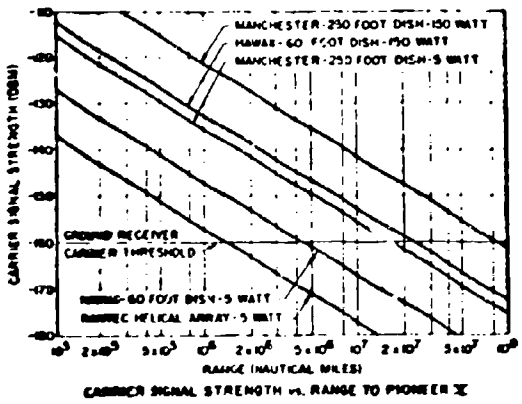
time is achieved.

For communication from earth to the payload a command receiver is employed. The command receiver is a transistorized double-conversion, phase-lock-loop receiver which produces a coherent output at 2/17 of the received frequency. It can be operated with either a 250-cps or a 40-cps bandwidth at threshold sensitivities of -130 dbm and -140 dbm, respectively. The receiver operates continuously and, since its bandwidth is considerably less than the expected long term drift of its nominal frequency plus frequency uncertainty of the received signal including doppler, it repeatedly sweeps over a range of 40 kc in wideband and 18 kc in narrowband searching for a carrier. Sweep period is 10 seconds for wideband and three minutes for narrowband. When the receiver acquires and locks on a signal from the earth, sweeping stops and the receiver can then accept any of eight possible commands, as follows:

1. Transmitters off
2. 5-watt transmitter on at 64 pps
3. 5-watt transmitter on at 8 pps
4. 5-watt transmitter on at 1 pps
5. Separate Stages III and IV
6. Receiver narrow band
7. 150-watt transmitter plates on
8. 150-watt transmitter filament on

Signals from the earth to the payload are transmitted by using a high-power carrier, phase-modulated with a 512-cps subcarrier. Amplitude modulation of the subcarrier with a coded train of 13 pulses provides the required information to the payload command decoder. The 13 pulses consist of one sync pulse, six information pulses and six parity check pulses.

The ground-to-air communication capability is shown in Figure 5 for three STL ground

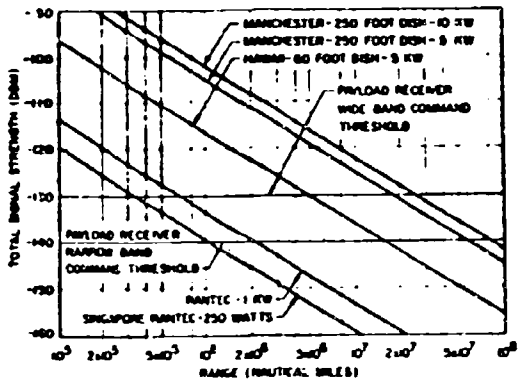


CARRIER SIGNAL STRENGTH vs. RANGE TO PIONEER X

Figure 5.

stations using 60- or 250-foot parabolic dishes and Rantec helical arrays. Ground transmitter powers of 250 watts, 1 kw, 5 kw, and 10 kw compatible with the ground station capabilities are also defined. A 3-db circular polarization loss is included on all curves; in Manchester a rotatable linear feed is available which can effectively improve the communication capability shown on the curve by 3 db.

The air-to-ground communication capability is shown in Figure 6. The curves are plotted



RECEIVER SIGNAL STRENGTH vs RANGE FROM PIONEER X

Figure 6.

for three typical ground station antennas with two payload transmitter powers. The ground receiver carrier threshold of -160 dbm has been achieved and the curves include a 3-db circular polarization loss which can, in the case of Manchester, be improved with the rotatable linear feed.

Actual measured payload radiated power levels were used in calculation. The radiated power was 3.15 watts phase modulated at one radian in the 5-watt case, or +35 dbm, and 85 watts in the 150-watt case, or +49.3 dbm. The carrier level has been reduced by 1.8 db in the curves to compensate for the power contained in the sidebands.

Tracking data has shown that the curves are within approximately 1 db of actuality.

The Telebit digital telemetry system accepts both analog and digital inputs from the various experiments and payload transducers for processing. This system converts the information into a word format shown in Figure 7 and sequentially reads out each word in succession. The output of the telemetry system is a biphase modulated 1024-cps subcarrier which is then directly phase modulated on the carrier.

The binary output of the Telebit system occurs at a synchronous rate and is composed of repeating sets of frames of words. For this payload, eight words per frame are used. One

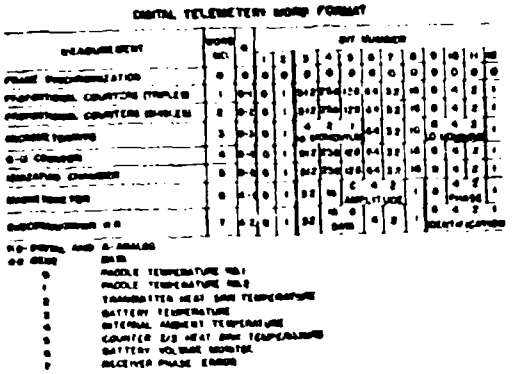


Figure 7.

word of each frame is used as a frame sync and is read out as all zeros, while the balance of the words are coded with the digital representation of the input information. Each word contains 12 pulses. The first two pulses (for information words) are always coded the same (zero, one) and define word sync. The other 10 pulses may take on any combination of binary values to represent a number from 0 to 1024.

A 12-bit combination binary counter and shift register, referred to as a shifting accumulator, is provided for each word. Pulses from the digital type of measurement are applied directly to the counting input of a shifting accumulator, while an analog input is applied to an analog-to-digital converter, whose output is then applied to a shifting accumulator. An electronic commutator running synchronously at the word rate gates 12 shift pulses to each shifting accumulator during one word interval each frame.

These shift pulses cause the information in the shifting accumulator to be delivered to the biphase modulator, and at the same time the output of the digital shifting accumulators is returned to the input so that after 12 shift pulses the state of the shifting accumulators is exactly as it began. The output of all of the shifting accumulators are connected together but since only one is shifting at a time no interference results.

The conversion of analog to digital information is done with a digital ramp and a voltage comparison circuit. In essence, conversion results from counting the number of steps in the ramp below the level of the analog input. The counting is actually done in a shifting accumulator, just as for digital experiments.

The biphase modulator accepts the pulses emerging sequentially from the shifting accumulators and produces a subcarrier whose phase shifts by 180 degrees each time a one is to be transmitted. This biphase modulated subcarrier is then delivered to the transmitter for phase modulation upon the carrier.

The pulses which cause the electronic commutator to step and the shift registers to shift originate in the programmer, and the programmer in turn receives its excitation from the master clock. The programmer is equipped such that application of an outside signal from the command decoder causes the pulse rate of the digital telemetry system to change.

4.2 Power Supply

Placing solar cells directly on the surface of the payload presented two difficulties: (1) the available surface area of the payload did not conveniently permit adequate solar energy interception; (2) temperature balance of the payload would have had to be too low for the internal instrumentation. An early decision was therefore made to locate the solar cells on paddles extending essentially radially from the payload. These paddles were folded within the nose fairing during launch and at second stage burnout after the fairing is jettisoned they sprung out and latched into place. A total of 4800 boron-diffused silicon solar cells are carried on the satellite. Because of attitude and spin considerations, however, only about 1200 of the solar cells are receiving solar energy at one time.

The payload storage battery consists of two packs of 14 nickel-cadmium cells each. Part of the power developed in the solar cells is immediately consumed in experiments and the receiver, and the charging rate of the battery is thus lessened by this amount. The 5-watt transmitter, drawing a large portion of its power from the batteries, has been operated approximately 10 per cent of the time; the 150-watt transmitter can be operated approximately 1 to 2 per cent of the time. The nominal power available from the solar cell conversion system is about 15 watts when the payload is near the earth and about 24 watts at perihelion.

The batteries operate nominally at 18 volts. Since this voltage is inadequate for all of the electronics, a series of static converters is employed to provide a variety of voltage levels. An under-voltage control is incorporated which automatically removes the transmitter load from the batteries in case battery discharge goes so far as to threaten to disable the receiver and thus prevent the payload from being commanded from the ground. In addition, the under-voltage control assures the conservation of battery lifetime by preventing deep battery discharges.

A thermal control is employed on the batteries which reduces the solar cell charging current to the batteries whenever the battery temperature rises above 120°F. This would occur during long periods of overcharge, when the batteries cannot be drained due to inability to command on the transmitters, and at the same time the internal impedance of either battery pack varies sufficiently to cause an unbalanced charge condition.

4.3 Temperature Control

The temperature of the payload is maintained within the desired range of 35 to 80°F by two methods. The heat generated by the transmitters and converters is absorbed by lithium heat sinks at the sites of this equipment. These heat sinks then radiate the heat to space and prevent operation of the transmitters and convertors from significantly increasing the general temperature of the payload. Moreover, the coating of the payload is carefully prepared so that the ratio of absorptivity to emissivity of solar energy of the payload shell varies with attitude in a manner carefully calculated to keep the temperature within the proper range as the payload moves in toward the sun.

Thin glass plates, 0.003-inch thick, on the solar cells, on the under surface of which an interference filter has been vacuum-deposited, serve to keep the temperature of the cells within oper ting range. The paddles are also properly coated between shingles of cells to keep temperature down.

All of these controls have operated satisfactorily. Soon after lift-off, for example, the solar cell paddles had stabilized at a temperature of approximately 25°F, just as had been anticipated. Gradually since then their temperature has increased, as the result of changing attitude with respect to the sun and of the increased intensity of solar energy as the sun is approached. Their temperature as of mid April was about 51°F. The maximum anticipated temperature of the paddles is about 65°F.

COMMUNICATION IN SPACE BY DEFLECTED SUNLIGHT

By: K. W. Otten, Communications Laboratory
Wright Air Development Division
Wright-Patterson AFB, Ohio

INTRODUCTION

Electronic equipments designed for long time operation in space vehicles such as satellites receive their primary power indirectly from the sun's radiation which is inefficiently converted into electrical power. In communication and telemetry equipment conversion of this electrical power into electro-magnetic radiation at one specific frequency results in further reduction of efficiency.

For long time observations and operations there is not always a need for immediate transmission of information. A delay between collection of the information and its transmission may be tolerated. In these cases a communication system using deflected sunlight as an information carrier might replace a radio communication system.

At a first glance space communication or telemetry systems using sunlight as an information carrier without conversion into electrical power seem attractive for the following reasons:

1. No lossy conversion of the solar radiation into electrical energy is necessary.

2. No lossy second conversion of the electrical energy into electrical-magnetic radiation is necessary.

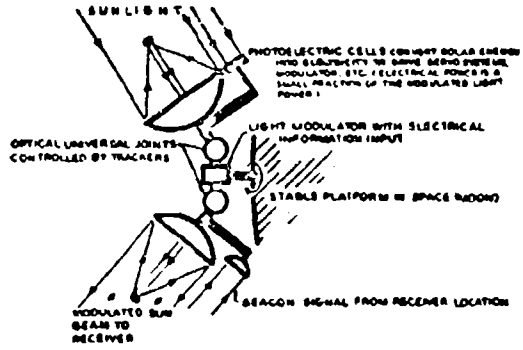
3. Simplicity and potential higher reliability of the equipment as compared with that of the radio frequency link.

4. High system gains are feasible due to the high directivity of optical systems.

As with most innovations, the advantages appear more lustrous before practical application has been attempted. The disadvantages, mostly hidden, are neglected or only unduly considered. The purpose of this study is to evaluate objectively the potential performance characteristics of sunlight communication links and to compare them with those of radio communication links.

DESCRIPTION OF THE SUNLIGHT COMMUNICATION LINK

Prior to the theoretical analysis of the sunlight communication link, the basic components of the system are described (Figure 1). At the transmitter station a large cassegrainian mirror system is aimed toward the sun to collect the solar energy and to concentrate the intercepted sun rays into a narrow beam with high light flux density. This beam, redirected by an optical universal joint, is passed through the light modulator which has a fixed position relative to the space vehicle or the transmitter location. The intensity or polarization modulated light beam is then redirected by a second optical universal joint, fed into another large cassegrainian mirror system which reradiates the modulated solar energy to the distant receiver.



SUNLIGHT COMMUNICATION TRANSMITTER TERMINAL

FIGURE 1

Power is required to keep the collector and radiator mirrors tracking the sun, and the receiver, respectively. The power for the tracking sensors and drivers together with the modulator power and the power required to operate the information processing equipment must be delivered from a solar battery. This power, however, can be a small fraction of the power being modulated.

The tracking of the sun generates no problems since the sun represents an excellent target under all conditions. The proper orientation of the radiator mirror system toward the receiver, in contrast, causes great difficulties. Proper orientation can be maintained either by employing a closed loop system or by having a knowledge of the receiver location with respect to the transmitter as a function of time. In Figure 1, the small receiver antenna mounted with its axis parallel to that of the radiator antenna represents the homing antenna for a closed loop system.

The receiver station consists of a cassegrainian collector mirror of possibly variable focal-strength to allow elimination of background noise after exact orientation toward the transmitter has been achieved. The photo-cathode of the multiplier phototube is located at the focal point of the collector mirror system. It converts varying light intensity into a varying electron current and amplifies it. Following the multiplier phototube stage, which represents the lowest noise combination of a photo detector and an amplifier, are noise filters and standard amplifiers. From here on the signal is processed like any other electrical signal.

DIFFERENCE BETWEEN LIGHT AND RADIO AS INFORMATION CARRIERS

Before the potential performance of the sunlight communication system can be compared to that of a radio communication system, the basic differences between the two types of carriers must be analyzed.

Electro-magnetic radiation as used in radio communication is generated by a controlled operation: selective amplification and feed-back with proper phasing. The generated signal can be specified as a periodic function of time. Amplitude-phase and amplitude-time relationships are described by defined functions. The signal is said to be coherent. This means that the signal can be identically duplicated by knowing the specifications of the function.

Electro-magnetic radiation at very high frequencies such as infrared or light frequencies is the result of an entirely different process. Externally stimulated nuclear particles in atoms or molecules change their energy levels. These energy level changes result in emission or absorption of specific energy quanta in form of photons. Only a limited number of energy level transitions are possible for each excited material. The energy radiated by one single quantum is directly proportional to the frequency at which the radiation occurs, the proportionality constant Plank's constant. The emission of the quanta, even if they all represent the same frequency, occurs randomly in time. The instants at which quanta are emitted are entirely out of any control. Due to the finite time required for an energy jump and to the random emission of quanta, the resulting signal occupies a certain frequency band instead of the frequency line, if we think of the visual spectrum display.

Radiation signals at very high frequencies can be described only by statistical parameters. An instantaneous relationship between phase (or time) and amplitude can not be established: Electro-magnetic radiation at light frequencies is incoherent.

The transition from the coherency of signals at low frequencies to the incoherency of signals at extremely high frequencies is gradual. With increasing frequencies, the problem of keeping the absolute bandwidth of the signal, or, stated another way, the absolute instability of the signal, as small as possible, becomes more and more difficult.

Lack of phase stability, or in our terms, lack of coherency of a signal reduces the scale of possible modulation schemes to that of various intensity modulation methods. But this drawback for communication is topped by another effect resulting from the randomness of the signal.

Electro-magnetic energy is radiated in quanta. Since the energy of each quantum increases proportionally with frequency, it can be seen that for the transmission of a certain amount of energy at low frequencies a large number of quanta is necessary but at high frequencies a much smaller number will suffice. Since the emission of the quanta occurs randomly rather than periodically, each individually detectable amount of energy must contain sufficient quanta so that the instantaneous variations of the quanta emission rate are not noticeably different from the average rate.

A numerical example may illustrate how the quantization noise component becomes important in electro-magnetic radiation signals at very high frequencies. Let us assume that the average number of quanta necessary to detect one bit of information with a satisfactory signal to noise ratio (it is assumed to be approximately 3 to 1), is 10, then the minimum power density n in watts per cycle bandwidth required at the input of an ideal photo detector would be at a function of frequency f:

$$n = 10\,hf$$

WITH $h = 6.625 \cdot 10^{-34}\ \mathrm{W/SEC^2}$

For a carrier frequency in the ultraviolet light range this amounts to approximately 10^{-18} W/cps. If we compare that value with the equivalent noise power of a radio receiver having a noise temperature of 300°K (10^{-21} W/sec) we realize the importance of the quantization noise.

In other words the amount of energy transmitted by a single quantum determines the absolute sensitivity of an ideal receiver. As an ideal receiver a receiver is specified which has 100% efficiency and which is free of internal noise.

THE ABSOLUTE MINIMUM RECEIVER SENSITIVITY

Let us assume that the receiver receives on an average basis one quantum per cycle bandwidth.

The instantaneous number of quanta per cycle bandwidth may vary between 0 and any multiple of 1. For this low number of quanta the input power can be considered a noise power and one can equate the energy of one quantum (photon) with the spectral noise power density n:

$$(2) \quad q = n = hf$$

This specific noise power density hf can also be expressed by an equivalent noise temperature:

$$n = hf = \kappa T_q$$

$$(3) \text{ OR } \quad T_q = \frac{hf}{\kappa}$$

If we assume that the receiver has an equivalent noise temperature T_r, the antenna an equivalent noise temperature T_a and the signal itself, due to its quantum nature an equivalent quantum noise temperature T_q, the noise power density at the receiver input is:

$$(4) \quad n = \kappa \left(T_a + T_r + T_q \right)$$

For an ideal receiver and ideal conditions, that is, equivalent receiver noise temperature and the equivalent antenna noise temperature both being zero, the noise power density becomes a function of the frequency (or wavelength of the light) only

$$(4a) \quad n = \kappa T_q = hf$$

The relationship between the noise power density n at the input of the receiver, or the equivalent noise temperature of the signal and the carrier frequency f is shown in Figure 2.

ABSOLUTE LIMIT IN RECEIVER SENSITIVITY

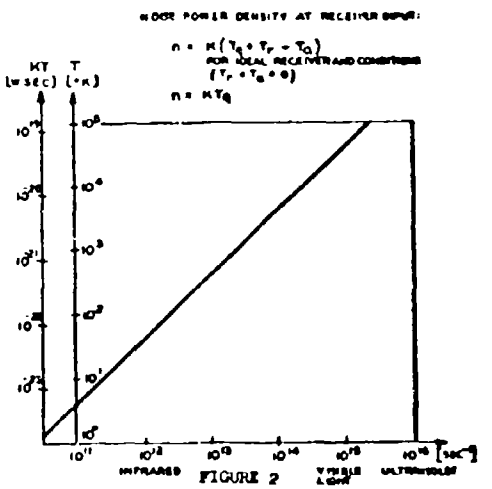


FIGURE 2

We see that for the receiver operating in the range of visible light ($f = 10^{15}$ cps) the absolute limit in sensitivity is described by an equivalent noise temperature of approximately 50,000°K. It must be noted that this noise figure is that of an ideal receiver. Practical photo-detectors as receivers have internal noise and therefore increase the equivalent noise temperature of the receiver considerably. In addition, the antenna usually "sees" not only the signal source but also background noise.

COMPARISON OF THE SUNLIGHT COMMUNICATION LINK WITH A RADIO COMMUNICATION LINK

Let us assume that an ideal transmitter radiates its entire power isotropically into space. We can then for comparison of various communication systems select a reference distance and compare the power required per cycle of bandwidth to obtain a desired signal-to-noise ratio at the receiver, in the absence of external noise sources.

We take a radio communication system with a receiver which has an equivalent noise temperature of 1500° Kelvin as our reference. The sunlight communication system will require 50 times as much power for the same information capacity, provided that it employs an ideal detector. For practical cases this power can be orders of magnitudes higher. A radio communication system powered by a solar cell is comparable with this ideal sunlight communication system, even if the overall efficiency of the photo-electric converter and the radio transmitter is as low as 2 percent. It must be noted that this is the theoretical limit even assuming a relatively high receiver noise temperature. Considering that practical photo multipliers have noise temperatures which are hundreds of times higher than this theoretical limit, the efficiency of a sun powered radio system can become a very small fraction of a percent to be competible with the hypothetical isotropically radiating sunlight system.

Even so the introduced equivalent isotropic radiator model is only of hypothetical value, it shows clearly that light is not suitable for systems incorporating wide beam width transmissions. It seems therefore advisable to employ a radio link for initial acquisition and tracking in a complete closed loop sunlight system, since the noise temperatures of radio detectors can be kept small.

The comparison becomes more advantageous for the sunlight communication system, if we consider the directional gain obtainable with it.

It can be shown that the combination of a parabolic collector with a parabolic reradiator as a transmitter in a sunlight communication link has a gain of 53 db over an isotropic source if the station is located at the earth's orbit and if the collector mirror is the only limiting aperture. By such an arrangement the entire collected solar energy is reradiated into a cone with an angle of 32 minutes or .0093 radians. This angle is identical to the angle under which the sun is seen from the earth.

To make the characteristics of the sunlight link system more understandable, a short analytical comparison with a radio link is presented.

To emphasize the basic differences and limitations the comparison is exercised for one way links only, neglecting the problems of acquisition, tracking and background noise elimination.

Radio links in general and sunlight links making full use of their directional gain are compared under the following assumptions:

1. The distance R between the stations is the same.

2. The effective antenna areas of the receiver antennas A_R are the same.

(2) $A_{R/RADIO} = A_{R/SUNLIGHT} = A_R$

3. The effective collector area A_C of the sunlight communications system is equal to its effective transmitter antenna area A_T.

(3) $A_C = A_{T/SUNLIGHT}$

As a result, the gain of the sunlight system transmitter is fixed (53 db), the beam width of the radiated beam is identical to the diversion angle under which the sun is seen from the earth's distance. (32 minutes of arc).

4. The effective transmitter antenna area of the radio link is equal to that of the sunlight collector and radiator.

(3a) $A_{T/RADIO} = A_{T/SUNLIGHT} = A_C$

5. The receiver sensitivity is not limited by background noise but by internal noise only.

6. The conversion in efficiency and the non-linear relationship between input signal-to-noise ratio and output signal-to-noise ratio of radio detectors and light detectors are neglected.

The systems are compared for unit signal to noise ratio at the input of the detector.

The noise power density at the input of an ideal receiver is:

(7) $n = KT_q = hf \left(\frac{W}{CPS}\right)$

$K = 1.38 \cdot 10^{-23} \frac{W}{sec \cdot {}^0K}$

$h = 6.625 \cdot 10^{-34} \frac{W}{sec^2}$

f = carrier frequency in cycles per second.

T = equivalent quantum noise temperature in degrees kelvin.

The noise power density at the input of the receiver is limited by the internal noise:

(8) $n = K(T_q + T_r)$

$\approx KT_r$ FOR $T_r >> T_q$

which is valid for frequencies below 10^{11} cps with T_R being the equivalent noise temperature of the receiver input.

At the receiver antenna the effective area A_R (m^2) which is located at the distance R in meters from the transmitter, the signal power P_R (W) becomes a function of the transmitter power P (W) and the transmitter antenna gain G:

(9) $P_R = \frac{P_T}{4\pi} G A_R \frac{1}{4\pi R^2}$

For the sunlight communications systems; the transmitter antenna gain is fixed by the choice of the ratio between collector and radiator (transmitter antenna) area: $\frac{A_C}{A_T}$

(10) $G_S = 1.85 \cdot 10^5$ OR 53 DB

The transmitter power P_{TS} is a product of the solar constant S and the effective area of the collector mirror A_C:

(11) $P_{TS} = S A_C$ then the signal power at the receiver is:

(12) $P_{RS} = \frac{S A_C}{4\pi} G_S A_R \frac{1}{4\pi R^2}$

For the radio communication link; the transmitter antenna gain is:

(13) $G = C_1 \frac{A_T}{\lambda^2}$

If A_T is the effective area of the radiating antenna, C is a constant determined by the type of the antenna and λ is the wavelength of the radio signal.

Due to the assumption $A_T = A_C$ we can write

(14) $G = C \frac{A_C}{\lambda^2}$ therefore

(15) $P_R = \frac{C}{4\pi} A_C \frac{P_T}{\lambda^2} A_R \frac{1}{4\pi R^2}$

We can now equate the noise power and the signal power available at the input of the receiver with the information bandwidth Δf (cps). We obtain for the sunlight system:

$$(16)\quad \Delta f K T_q = \frac{S G_3}{4\pi} A_c A_R \frac{1}{4\pi R^2}$$

and for the radio communication systems

$$(17)\quad \Delta f K T_r = \frac{c P_T}{4\pi \lambda^2} A_c A_R \frac{1}{4\pi R^2}$$

We can eliminate the factors and parameters which are identical to both links according to the assumptions made above by dividing (16) by (17):

$$(18)\quad \frac{T_q}{T_r} = \frac{S G_3}{c} \cdot \frac{\lambda^2}{P_T} \qquad \text{OR}$$

$$(19)\quad P_T = \frac{S G_3}{c} \cdot \frac{T_r}{T_q} \lambda^2$$

This expression describes the relationship between the equivalent noise temperatures of the radio receiver T_r, of the sunlight signal T_q, and the power of a radio transmitter P_T with carrier wave length λ which is required to obtain the same performance as the sunlight link. While equation (19) compares the ideal sun light links (no internal noise at the detector) with practical radio links, it can be extended to technically feasible sunlight links by replacing the noise temperature of the signal T_q by the much higher noise temperature of the multiplier phototube T_q':

$$(20)\quad P_T = \frac{S G_3}{c} \cdot \frac{T_r}{T_q'} \lambda^2$$

In Figure 3 this relationship is presented in graphical form. For practical cases $\frac{T_r}{T_q'}$ is in the order of 10^{-4}, but can be smaller for very low radio receiver noise temperatures of higher for high noise temperatures.

It must be noted that the directional properties of the radio link cannot be considered for this comparison. While the beam width of the sunlight transmitters is fixed, it can be varied at the light receivor antenna without changing the effective area. In contrast, for a radio antenna, the beamwidth for any selected carrier frequency is determined by the effective area.

It must be noted that this comparison holds independently of the information bandwidth and the range of the link as long as the antenna areas are as specified.

For a valid comparison between radio and sunlight communication systems, we must assume that the radio communication system will be powered by a solar battery.

If we assumed that the required power P_T of the radio transmitter will be delivered from a solar battery with an efficiency η and with an effective collector area A_C, we obtain:

$$(21)\quad P_T = \eta S A_C$$

If we insert this value in equation (17) and if we make the additional assumption that the collector area of the solar battery shall be identical to the transmitter antenna area and to the collector area of the sunlight communication system, then we can write:

$$(22)\quad \Delta f K T = \frac{c \eta S}{4\pi \lambda^2} A_C^2 A_R \frac{1}{4\pi R^2}$$

By elimination of the identical factors in equation (16) and (22) an expression is formed which allows the determination of the specific size of a collector and transmitter antenna for which radio and sunlight link have same information capacities.

$$(23)\quad A_C = \frac{G_3}{c\eta} \cdot \frac{T_r}{T_q'} \lambda^2$$

This function is graphically shown in Figure 4, with $\frac{T_r}{T_q'}$ as a parameter. ($\eta = 0.1$ corresponds to a combined converter-transmitter efficiency of approximately 1 percent.

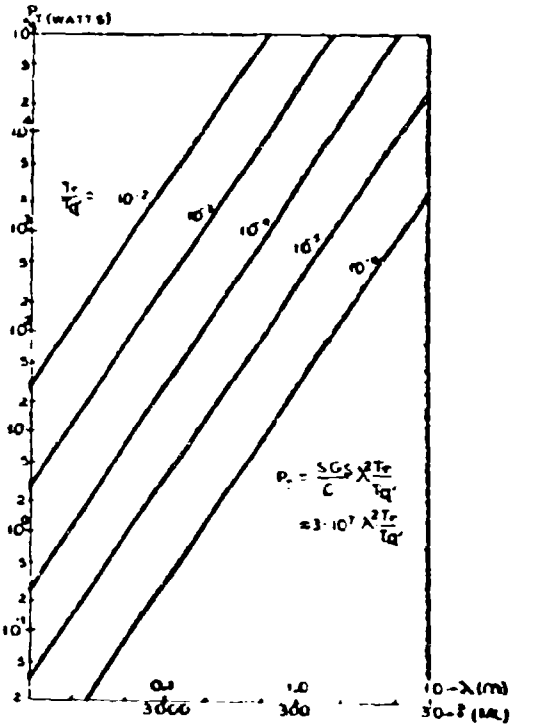


FIGURE 3

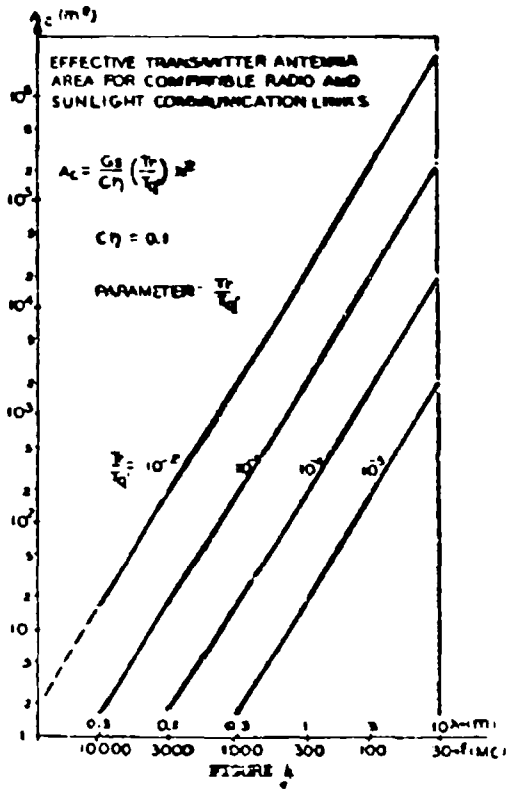


FIGURE 4

A comparison of sunlight and radio links on the basis of efficiency is only one way of evaluating expected performances. It permits one to estimate how a link could perform under somewhat idealized conditions. For practical purposes additional factors must be considered which cannot be expressed clearly in analytical form. Some of these factors are listed below:

1. The background noise seen by the light receiver mirror system depends largely on the focal length of the system and the location of the transmitter in space.

2. Whenever the sunlight system can be made compatible with a radio system on the basis of efficiency, the requirements for accuracy in directing the antennas because more severe for the sunlight mirrors than for the radio antennas.

3. The power required for the modulation in the sunlight transmitter amounts to a much smaller fraction of the modulated power than in the radio case.

4. The electronic equipment needed for the modulator in the sunlight transmitter can be built with a small number of components with low power dissipation. It may permit longer lifetime and higher reliability.

5. The requirements for tracking the sun are more severe for the sunlight system than for a radio system powered by a solar battery.

Keeping the transmitter antenna oriented toward the remote receiver is no serious problem, provided the contact has been established. To establish contact with a receiver of unknown location, however, sunlight techniques are not suitable. A closed loop system could consist of a sunlight link for the transmission of the information and radio link for the control of the information flow and the proper orientation of the antennas.

CONCLUSIONS

Even though radio communication links are more efficient than a sunlight communication system, due to the high noise temperatures of signals at light frequencies, there might be a number of advantageous applications of the sunlight links in space.

Due to the fixed gain of the sunlight transmitter which is independent of the size of the collector and reradiator mirror, the sunlight system can be made superior to a comparable radio system when only physically small antennas must be used, and if extremely high frequencies are not desirable for other reasons such as lifetime and reliability of components. This superiority in information handling efficiency is of practical value only, if the link is intended for long time operation, if the contact between the stations can easily be established and maintained, and if the information need not to be transmitted instantaneously.

APPLICATION OF INERTIAL TECHNIQUES TO INTERPLANETARY NAVIGATION

Milton J. Minneman
Republic Aviation Corporation

ABSTRACT

Inertial navigation techniques, at present used for the indication of the position of a vehicle near the surface of the earth may also be used for interplanetary navigation. Many methods have been suggested for interplanetary navigation, however inertial techniques are of interest because they provide acceleration and velocity as well as position. Also, since inertial techniques have been well developed in the last decade they should be exploited for interplanetary in addition to terrestrial navigation.

An interplanetary navigation system would employ an orthogonal array of three accelerometers. The output of the array, considered as a vector, is the specific thrust which equals the acceleration of the vehicle minus the specific gravitational force on the vehicle due to all bodies in space. If the position of the vehicle is known relative to the various bodies in space at one point in time, the value of the specific gravitation can be found. It can then be added to the accelerometer output obtaining the acceleration. If in addition the initial velocity is known, the acceleration can be integrated twice, with respect to time to give a new position of the vehicle at the next interval of time. From this new position, the new specific gravitation can be computed and added to the accelerometer output, etc., with the cycle repeated indefinitely. The necessary alignment of the accelerometers may be obtained by means of a combination of gyroscopes and optical sensors.

Using an over-simplified model of the solar system, the performance and accuracy of the system on simulated missions to Mars and Venus, using a low thrust engine, have been investigated analytically, and with the aid of analog and digital computers. It is demonstrated that the system can be used for navigation with an accuracy of approximately 0.01% of the distance traveled.

This amounts to about three (3) Venus radii for an Earth to Venus mission and nine (9) Mars radii for an Earth to Mars mission. The extension of the analysis to more realistic models of the solar system is described, and a contemplated complete system configuration is outlined.

SYSTEM DESCRIPTION

Inertial navigation techniques, at present used for the indication of the position of a vehicle near the surface of the earth may also be used for interplanetary navigation. Many methods have been suggested for interplanetary navigation, but inertial techniques are of interest because they provide acceleration and velocity as well as position. Additionally, inertial techniques have been well developed in the last decade and the technology should be further exploited.

In a terrestrial inertial navigation system, accelerometers sense the components of acceleration of the vehicle in orthogonal directions. The accelerometer outputs may be integrated once, yielding the vehicle's velocity, and again to yield position. A Schuler loop is used to maintain the sensing axes of the accelerometers perpendicular to the earth's gravitational field.

Attempting to apply these techniques to interplanetary navigation, various difficulties are encountered. It is well known (Ref. 1) that an accelerometer reads the non-field force per unit mass (specific force) applied to the vehicle in which the accelerometer's case is rigidly mounted. The gravitational field force, acting equally on the seismic mass and its case produces no relative displacement of the mass and therefore no indication. Thus, in the case of an interplanetary space vehicle, where the vehicle's acceleration is determined by the resultant of the specific applied thrust and the specific gravitational field force, only a portion of the acceleration is measured. Expressing the above mathematically

$$a = T/M = \ddot{R} - g \qquad (1)$$

where a is the reading of the accelerometer

T is the applied thrust

M is the mass of the vehicle

$\ddot{R}$ is the total acceleration of the vehicle

g is the specific gravitational force at the position of the space vehicle

Thus it may be seen that the accelerometer's readings give the vehicle's acceleration due to all causes less the specific gravitational force. It is apparent that if g could be obtained, it could be added to "a" yielding $\ddot{R}$, the actual acceleration of the vehicle.

The specific gravitation is equal to the resultant of the gravitational forces due to the sun, earth, other planets, planetary satellites, etc. The effect of man-made bodies, and the stars are negligible as will be shown later. The expression for specific gravitation is

$$g = \sum_m g_m = \sum_m \frac{M_m G_o (R_m - R)}{|R_m - R|^3} \qquad (2)$$

where

g_m is the specific force due to the m^{th} body

M_m is the mass of the m^{th} body

G_o is the universal gravitation constant

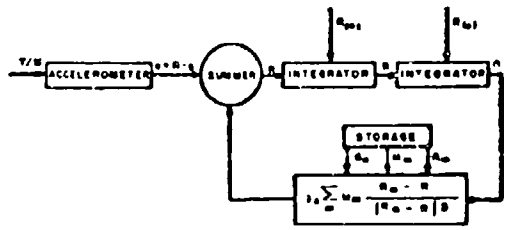
R_m is the position of the m^{th} body

R is the position of interest

The quantity M_m is known for each body of interest and R_m the position of the body may be found from an ephemeris as a function of time. Thus, if time and the vehicle's position are known g may be computed. Using this in equation (1), $\ddot{R}$ can be computed and hence $\dot{R}$ and R can be computed. R, in turn, can be used to compute g. If, in addition, R and $\dot{R}$ of the vehicle are known at some time, t, say t = O, then there is available all the information required for the continuous computation of R, $\dot{R}$, and $\ddot{R}$.

Specifically, starting at time, t = O, the known value of R is used to compute g from equation (1). g and the accelerometer output "a" are combined to compute $\ddot{R}$. $\ddot{R}$ is integrated with respect to time giving an increment in velocity which is added to the known value of $\dot{R}$ at t = O to give a new value of $\dot{R}$ at the new $t=t_1$ (the time of integration). Simultaneously, $\dot{R}$ is integrated with respect to time giving an increment in position which is added to the known value of R at t=O to give a new value of R at $t = t_1$. This process is repeated indefinitely.

The computation loop is shown in Figure 1. The diagram shows the ideal case in which the computed and stored values are assumed correct. Possible errors and their effects are discussed later.



BASIC NAVIGATION LOOP

FIGURE 1

In the above, all quantities must be measured or operated on separately for each component. In an actual computer, three loops similar to the above would be used, one for each of the axes of an orthogonal coordinate system such as cylindrical or rectangular.

The individual loops are inter-related by the necessity of computing the distance from the attracting bodies to the vehicle, and by acceleration terms arising from the rotation of the accelerometers.

Regardless of the choice of coordinate system, the direction of the accelerometer axes must always be known, preferable along the axis directions of the chosen system. Some type of stable platform for the accelerometers is required. It is contemplated that such stabilization would be accomplished by a combination of short term gyroscopic stabilization and long term corrections from solar, planetary or stellar radiation.

GRAVITATION, THRUST AND TRAJECTORIES

It was stated in the previous section that an accelerometer reads only the portion of the acceleration of a space vehicle due to thrust and not that due to gravitation. Since the navigation system must add the gravitational term to the accelerometer's output, it is instructive to determine the specific gravitational force on various bodies. In considering this, a specific thrust of 10^{-3} meter/sec^2, developed by an electrical propulsion system, is used as a reference. Gravitational terms as low as 0.01% of this value or 10^{-7} meters/sec^2 would be significant.

From these criteria and appropriate gravitational computations the following conclusions may be reached:

In general, three bodies must be considered; the sun must always be considered, near the earth and moon they must also be considered. As the vehicle approaches the destination planet, its gravitational field must be considered.

It was previously stated that an acceleration of 10^{-3} meter/sec^2 would be used as a reference. This is based on the assumption that the guidance system described herein will provide its greatest utility in a propelled rather than a ballistic vehicle. A vehicle suitable for a mission to Mars using a plasma propulsion system is described in Reference 2. A study has been made of trajectories using such a propulsion system (Reference 3); it is indicated that a typical mission would be accomplished in three phases:

a) An outward spiral around the earth

b) A transfer from the heliocentric orbit of the earth to that of the destination planet

c) An inward spiral around the destination planet

An equiangular spiral is assumed as a reasonable trajectory for such a mission; it is shown that the radial and transverse components of total acceleration, and of gravity are all of the same order of magnitude. Thus it appears reasonable to use the figure of 10^{-3} meter/sec^2 as a reference figure for comparison. For subsequent analyses, trajectories based on the theory expressed in Reference 3 will be used.

GENERAL BEHAVIOR OF LOOP

It is instructive to examine the behavior of the computer loop of the navigation system under ultra-simplified conditions, namely; in a vehicle traveling along a straight line, either directly toward or directly away from the sun. Only the effect of the sun is considered. Conclusions should not be drawn from this examination, but some trends may be seen.

Figure 1 defines the loop as

$$\ddot{R} = (\ddot{R} - g) + \frac{M_m G_o}{(R_m - R)^2} \qquad (3)$$

In the simplified one dimensional case under consideration this becomes

$$\ddot{X}_o = \ddot{X} + \frac{K}{X^2} - \frac{K + \alpha}{X_c^2} + \ddot{e}_o \qquad (4)$$

where

X = the actual position of the vehicle on the X axis

X_c = the computed position of the vehicle on the X axis

K = the actual value of $M_s G_o$

M_s = Mass of the sun

$K + \alpha$ = the value of $M_s G_o$ used the system

$\ddot{e}_o$ = the error in the accelerometer

Defining the navigation system error as

$$e = X_c - X \tag{5}$$

so that

$$\ddot{e} = \ddot{X}_c - \ddot{X} \tag{6}$$

equation 2 may be rewritten as

$$\ddot{e} = \ddot{e}_o + \frac{K}{X^2} - \frac{K + \alpha}{(X + e)^2} \tag{7}$$

The quantity $K = M_s G_o$ is related to the astronomical unit X_E by virtue of Kepler's law

$$\frac{K}{X_E^3} = \frac{4\pi^2}{y^2} = \text{a constant} \tag{8}$$

where y is the length of the year. Therefore α, the error in assumption of K may be expressed in terms of e_E, the error in assumption of X_E, as follows:

$$\alpha = \frac{3K}{X_E} e_E \tag{9}$$

which may be substituted in equation (3) yielding

$$\ddot{e} = \ddot{e}_o + K\left[\frac{1}{X^2} - \frac{1}{(X + e)^2}\right] - \frac{3K}{X_o(X + e)^2} e_E \tag{10}$$

This equation is a non-linear differential equation for which there is no ready solution. However, if $e \ll X$, and X is assumed approximately constant, an approximate analytical solution may be obtained as follows:

$$e = \frac{\ddot{e}_o}{\beta^2}(\cosh \beta t - 1) + \frac{\dot{e}_o}{\beta} \sinh \beta t + e_o \cosh \beta t + e_E \left[1.5 \frac{X}{X_E} - \left(1.5 \frac{X}{X_E} - 1\right) \cosh \beta t\right] \tag{11}$$

where $\dot{e}_o$ is the error in determination of the initial velocity

e_o is the error in determination of the initial position

$$\beta^2 = \frac{2K}{X^3} \tag{12}$$

In this usage e_o represents the error with respect to the earth's position, and e_E the error in the earth's position; i.e. the error in the astronomical unit.

The effects of various types of error sources may now be seen. Accelerometer errors and initial velocity errors cause a navigation error which gradually increase from zero with time. Navigation error due to initial position error increases gradually from its value at zero time. Errors resulting from the unknown value of e_E cause an error whose nature is not immediately apparent but the possibility of it being a small value is apparent.

To evaluate this equation it is necessary to consider a particular mission, and particular values of the sources of errors. This has been accomplished from both equation 11 and by computer solution of the exact equation. This evaluation showed that although the navigation error increased with time, it remained within reasonable bounds. In particular the error in navigation due to the uncertainty of the astronomical unit remains within quite tolerable limits.

Accordingly solution of the problem in the more realistic two-dimensional case was undertaken.

TWO-DIMENSIONAL SYSTEM

In this section a model will be established of a vehicle traveling in a plane containing the sun. The effects of the planets are ignored.

The actual trajectory of a vehicle in an Earth to Mars or Earth to Venus mission will be very nearly in a plane containing the sun, so that the assumption of the vehicle traveling in such a plane is reasonable. However, since the planets have significant effect on the trajectory during certain positions of the space vehicle, the results still cannot be considered conclusive but merely indicative.

The orientation in space of the orthogonal pair of accelerometers can be arbitrarily chosen. However, two logical possibilities are:

A. Accelerometer array such that one unit is always along a line from the center of the sun.

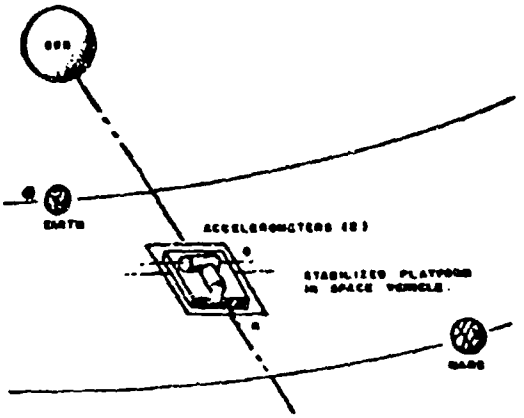
B. Accelerometer array fixed in inertial space.

In either array, both accelerometers are aligned in the plane of vehicle motion. An analysis of the sun oriented accelerometer array will now be made.

Sun-Oriented Accelerometer Array

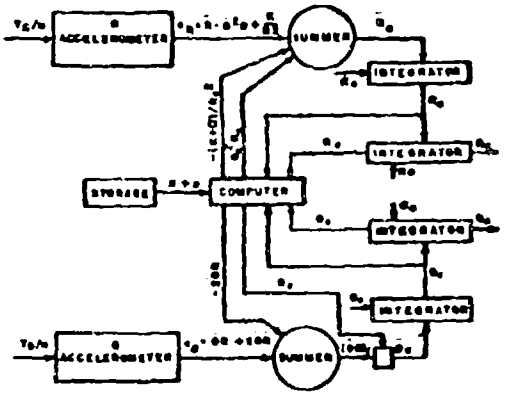
Figure 2 illustrates the space and platform geometry of such a system. This will hereafter be referred to as an R-THETA (R-θ) system. Figure 3

shows the accelerometer-integrator-computer elements of the system in block form. R and θ are used for coordinates, and the subscript c is used to distinguish computed quantities from the actual quantities.



SPACE AND PLATFORM GEOMETRY

FIGURE 2



R-θ NAVIGATION SYSTEM

FIGURE 3

The "R" accelerometer reads the sum of the radial acceleration, R, the centripetal acceleration, $-\dot{\theta}^2 R$, less the gravitation, $-K/R^2$. To obtain R_c, $\ddot{R}_c$ must be integrated twice. Therefore, it is necessary to subtract the centripetal acceleration from, and add the gravitational term to, the output of the accelerometer.

The "θ" accelerometer measures the transverse acceleration, $\ddot{\theta} R$, and the coriolis acceleration $2\dot{\theta}\dot{R}$ (no gravitation is measured since this accelerometer is assumed orthogonal to the sun's field). The coriolis acceleration must be subtracted from the accelerometer output and the result multiplied by 1/R. Two integrations then yield the angular position.

The "R" and "θ" computational channels are interrelated by the centripetal and coriolis corrections. Performance differences are thereby expected from the one-dimensional case.

The equations for the R-θ case may be written by inspection from Figure 3.

$$R_c = \iint \left[\ddot{R} - \dot{\theta}^2 R + \frac{K}{R^2} + \dot{\theta}_c^2 R_c - \frac{K+\alpha}{R_c^2} \right] dt\,dt \qquad (13)$$

$$\theta_c = \iint \frac{1}{R_c} \left[\ddot{\theta} R + 2\dot{\theta}\dot{R} - 2\dot{\theta}_c \dot{R}_c \right] dt\,dt \qquad (14)$$

To allow for accelerometer errors the quantities $\ddot{\alpha}$ and $\ddot{k}$ respectively should be added to each bracketed quantity.

Equations (1) and (2) are readily convertible to the form necessary for error analysis by substituting the following terms in them

$$e = R_c - R \qquad (15)$$

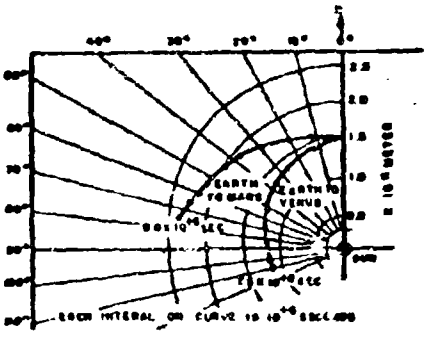
$$\delta = \theta_c - \theta \qquad (16)$$

This yields

$$\ddot{e} = \dot{\theta}^2 e + 2\dot{\theta}\dot{\delta}R + 2\dot{\theta}\dot{\delta}e + \dot{\delta}^2 R + \dot{\delta}^2 e - \frac{\alpha}{R^2} - \frac{K+\alpha}{(R+e)^2}\left[\frac{\alpha}{K} + \frac{e^2}{R^2} \right] + \ddot{\epsilon}_e \qquad (17)$$

$$\ddot{\delta} = \frac{1}{(R+e)} \left[-\ddot{\theta} e - 2\dot{\theta}\dot{e} - 2\dot{R}\dot{\delta} - 2\dot{\delta}\dot{e} + \ddot{k}_\theta \right] \qquad (18)$$

In order to obtain machine solutions of the above equations, it was necessary to select typical trajectories. Again, using the method of Reference (3), a pair of trajectories (one to Mars and one to Venus) was selected, both suitable for use with an electric propulsion system. Figure 4 illustrates the trajectories used.



TWO DIMENSIONAL TRAJECTORIES

FIGURE 4

The trajectory constants used in plotting these curves are given in Table I.

Trajectory Constant	MARS	VENUS
R (0)	$+1.496 \times 10^{+11}$	$+1.496 \times 10^{+11}$
$\dot{R}$ (0)	$+1.0 \times 10^{+4}$	$-0.5 \times 10^{+4}$
$\ddot{R}$	-0.25×10^{-3}	-0.125×10^{-3}
Θ (0)	0	0
$\dot{\Theta}$ (0)	$+1.97 \times 10^{-7}$	$+1.97 \times 10^{-7}$
$R\ddot{\Theta}$	-1.96×10^{-3}	$+2.14 \times 10^{-3}$
T	$+9.00 \times 10^{+6}$	$+7.50 \times 10^{+6}$

All quantities are MKS units

TWO-DIMENSIONAL TRAJECTORY CONSTANTS

TABLE I

In order to obtain numerical data, standard values of error sources have been chosen as follows (in MKS units):

Error Source	Magnitude
e_Θ	$1 \times 10^{+4}$
$\dot{e}_\Theta$	1
$\bar{e}_\Theta$	10^{-7}
δ_Θ	10^{-8}
$\dot{\delta}_\Theta$	10^{-12}
$\bar{k}_\Theta$	10^{-7}
e_E	30×10^{6}
a	$8 \times 10^{+16}$

MAGNITUDES OF ERROR SOURCES

TABLE II

In general these were selected by assuming the errors in position or instruments were approximately 0.01% of the quantity being measured. The value of a used is that corresponding to the value of e_E shown.

In Figure 5 on the following page there are plotted curves of e versus t for the several error sources e_Θ, $\dot{e}_\Theta$, $\bar{e}_\Theta$, δ_Θ, $\dot{\delta}_\Theta$, k_Θ, e_E for the Earth to Venus case. The curves for the Earth to Mars case are similar.

Examination of these curves shows most are of the general exponential nature previously noted for the one-dimensional case. However, it is noted that several show distinctly oscillatory characters. Examination of the $\xi(t)$ and $\delta(t)$ data (not included in this report) shows a considerable degree of oscillatory behavior indicating that the very long term trends of many of the curves are sinusoidal. We note that two error sources, e_E and $\dot{e}_\Theta$, predominate as for the one-dimensional case.

To obtain an overall estimate of the total navigational error it is assumed that each error source is independent of any other and may have either a plus or minus sign. The root sum square of all errors is then obtained. This has been carried out for $t = T$ for the Mars and Venus Mission with the results shown in Table III.

Mission	Radial Error	Transverse Error	Combined Error
Mars	$21.2 \times 10^{+6}$ m	$19.4 \times 10^{+6}$ m	$29.6 \times 10^{+6}$ m
Venus	$9.1 \times 10^{+6}$	$15.3 \times 10^{+6}$ m	$17.6 \times 10^{+6}$ m

ERROR SUMMARY - RΘ SYSTEM

TABLE III

In making the above error computation one other factor, not previously mentioned, has been taken into account. This relates to the uncertainty in knowledge of the semi-major axis of a planet's orbit which in turn is related to the uncertainty in the Astronomical Unit. Since the ratio of these semi-major axes to that of the earth's is well known, an error in assumption of the Astronomical Unit will produce an error in the radius of any planet's orbit by a proportional amount. To obtain a proper interpretation of the system accuracy, this error in planetary position must be subtracted from the error generated by the inertial navigation system. This has been done in the proper statistical manner in arriving at Table III.

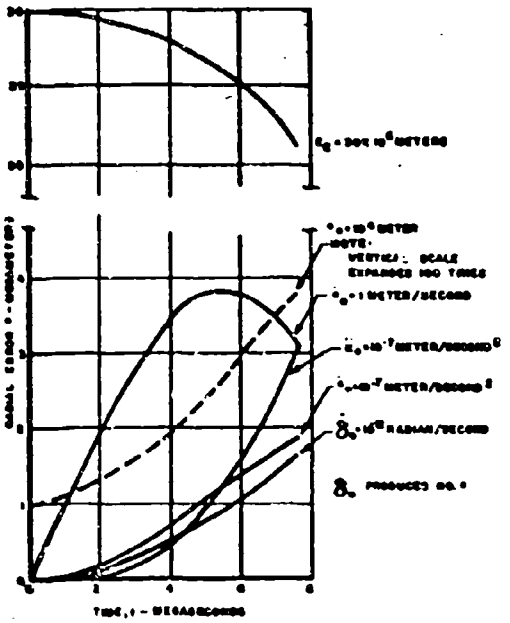
For purposes of comparison, the length of each trajectory is given below:

Earth to Mars $24 \times 10^{+10}$ m

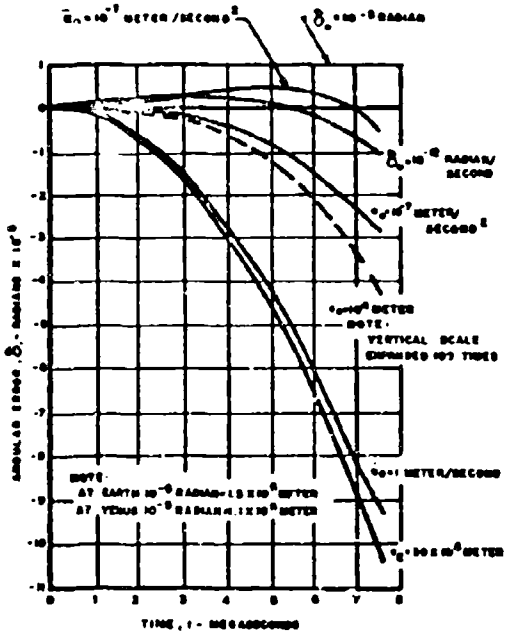
Earth to Venus $22 \times 10^{+10}$ m

It is thus seen that the error is in the order of 0.01% of the mission length. In setting up the original estimated values for error sources a figure of approximately 0.01% of the quantity being measured was selected. Thus, it is seen that the overall system performance is consistent with the accuracy of the instruments and measurements. This ratio of navigation accuracy to both instrument and initial condition accuracies may possibly be used as a figure of merit for navigation systems.

Another method of evaluating the accuracy is to note the relationship between the errors and the planet radii. The radius of Mars is approximately $3.3 \times 10^{+6}$ meters. The miss distance is thus approximately $29.2 \times 10^{6} / 3.3 \times 10^{6} = 9$ Mars radii. Estimates have been made that satisfactory homing on a planet can be made with miss distances exceeding 50 planet radii. For Venus with its larger radius of $6.1 \times 10^{+6}$ meters, a similar calculation shows a miss distance of approximately 3 Venus radii, an even more favorable figure.



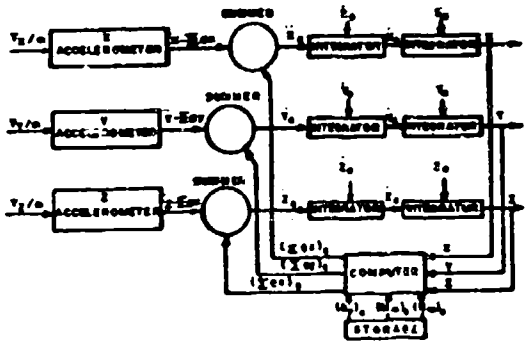
RADIAL ERROR
TWO-DIMENSIONAL R-θ SYSTEM
EARTH TO VENUS TRAJECTORY
FIGURE 5



ANGULAR ERROR
TWO-DIMENSIONAL R-θ SYSTEM
EARTH TO VENUS TRAJECTORY
FIGURE 6

MULTI-BODY THREE-DIMENSIONAL SYSTEM

Extension of the system to the multi-body, three dimensional case is now considered. The techniques described in the previous sections are directly applicable to establishing the system configuration. Figure 7 is a block diagram of this system with the orientation of the accelerometers fixed in inertial space.



THREE-DIMENSIONAL, MULTI-BODY, NAVIGATION SYSTEM
FIGURE 7

The equations for the system may be written generally as

$$x_c = \iint \left[\ddot{x} - \sum g_x + (\sum g_x)_o + \ddot{m}_o \right] dt\,dt \qquad (19)$$

etc.

Consideration must be given to the choice of coordinate systems. Conversion from one set of coordinates to another is a standard procedure and introduces no difficulties, so that different sets of coordinates may be used for separate purposes or at different configurations. To compute positions from the accelerometer readings it may be desirable to use geocentric coordinates near the earth and heliocentric coordinates in interplanetary space. For vehicle control purposes it may be desirable to compare the actual or computed trajectory with the scheduled trajectory in ecliptic equinoctical axes, and simultaneously to compute the trajectory in a set of axes whose X-Y plane is the approximate plane of the scheduled trajectory.

Consideration must also be given to the method of obtaining the positions of the attracting bodies for use in the equations. There exists several possibilities: Ephemeris data may be stored for planetary positions at stated time intervals, interpolation being used within the time intervals. The orbits of the planets can be described in the form of suitable series which would yield planet position coordinates at all times. Also since a space vehicle is only under the influence of any particular planet for perhaps 30 days, it may be possible to represent the planet's position by a simple equation.

Several other problems must be considered. The solar system is not inertial, but is rotating about the galaxy center. If necessary, a more generalized set of equations, which do not assume that the sun is fixed inertially, can be established, and the position of the vehicle with respect to the sun can be obtained even though the sun is not inertially fixed.

ACCELEROMETER AXES STABILIZATION

Brief consideration of the problem of axis stabilization may be made. The order of magnitude of the necessary alignment accuracy may be determined.

From Figure 8, it may be determined that an approximate expression for $\alpha(t)$ versus $\ddot{s}_\theta$ for the Venus trajectory is

$$\alpha(t) = 0.61(0.5\,\ddot{s}_\theta t^2) \qquad (20)$$

Now if the radial axis is misaligned by β radians, the radial accelerometer will give a reading in error equal to the transverse thrust times sin β. The effect of this is the same as an error in accelerometer. This may be substituted in the above equation yielding a relationship between allowable misalignment and resultant error. Using the Venus mission, an average thrust of 10^{-3} meters/sec^2, an allowable resultant error of 20×10^6 meters, an estimate of the allowable misalignment is obtained as $\beta = 1.17 \times 10^{-3}$ radians or 4 minutes of arc.

The above computation should be considered only as an indication of the nature of the problem. Since the error due to accelerometer misalignment depends on a variable thrust a dynamic analysis is required.

In terrestrial inertial navigation systems, gyroscopes are used for accelerometer axes stabilization. Gyroscope drift accumulation indicates the need for some in transit alignment correction for which optical or radiation sensing methods may be used. If such correction is made once every 24 hours, a maximum drift rate of 0.0028 degrees per hour would be permissible.

SYSTEM CONFIGURATION

A complete navigation system requires an array of aligned accelerometers, a computer, and an accurate clock.

Since the duration of a mission is of the order of 10^7 seconds and clocks with an accuracy of better than one part in 10^7 are readily available, the clock imposes no restriction on the system.

The digital computer used can be quite slow. The three accelerometer readings and the time of reading can be put into a buffer storage. Hence the computer speed requirements are determined by the interval of integration and by the possible need of navigation information within any such interval. From the digital computer runs made so far, large integration intervals can be used.

Preliminary analysis indicates that the alignment of the accelerometers must be accurate to within four (4) minutes of arc and each accelerometer must have about .01% accuracy. If the navigation system is used in a vehicle in which the thrust is applied intermittently, the disconnecting the accelerometers' outputs from the computer whenever there is no thrust could eliminate a substantial amount of the error due to the accelerometers.

The optical system used for alignment of the accelerometers may also provide limited navigation information directly. A combined optical, inertial navigation system provides complimentary information (lines of position, dynamic quantities respectively) which may be of advantage. Further the accuracy requirements of components of each system individually may be unrealizable whereas these are eased for a combined system. However, before a combined system is considered the characteristics of the inertial system must be investigated in greater detail.

SUMMARY

The work performed to date indicates that a reasonably accurate inertial navigation system is feasible. Overall accuracy of the order of the accuracy of the instruments and initial data is obtainable. This is true even in the face of the sizeable error in the value of the astronomical unit. With 0.01% initial data and instrument accuracy and no accelerometer alignment error, navigation errors are of the order of three (3) and nine (9) target planet radii in Earth to Venus and Earth to Mars missions respectively. If the accelerometer alignment can be kept within 4 minutes, the navigation error will increase less than 50%.

A word of caution is in order. The effect of the planetary gravitation terms has not been considered and this may have a considerable effect on the system. However, the results to date indicate the desirability of continued analysis.

ACKNOWLEDGEMENTS

The author wishes to express his appreciation for the assistance offered by various personnel who performed various portions of the analytical and computational work. Particular thanks are due to Dr. J. V. Hughes for furnishing standard trajectories, Messrs. J. Port and E. Schatz for analytical studies, Mr. J. Fussell for analog computer studies, and to Messrs. S. Broder, L. Lefton and Miss E. Smith for digital computer studies.

REFERENCES

1. Inertial Guidance
by Walter Wrigley, Roger B. Woodbury and John Hovorka; January 31, 1957, Sherman M. Fairchild Fund, Paper No FF-16.

2. An Experimental Plasma Propulsion System
by Milton J. Minneman; 1959 Conference Proceedings, 3rd National Convention on Military Electronics, pp 167.

3. Trajectories for Continuously Applied Propulsive Thrust
by J. V. Hughes and G. N. Nomicos; Proceedings of the Fourth Symposium on Ballistic Missile and Space Technology, 1959; to be published 1960.

HIGH PERFORMANCE CAMERA TUBES*

By: Dr. E. Gray, RCA Laboratories

The resolution performance which has come to be accepted in common practice for television camera tubes may be much poorer than the ultimate capability of these image transducers. In order to explore the limits of camera tube resolving capability and to produce tubes which can transmit more informational detail, an investigation has been undertaken of the imaging processes of which the tube is constituted. Further development of resolution is useful to fill the bandwidth of conventional transmission rates; for reconnaissance, it is a necessity.

This work has been directed toward tubes of the image orthicon class. The results are applicable to other camera tubes. The purposes of this discussion are to outline the methods of attack for systematic resolution improvement and to present experimentally derived information regarding some of the imaging processes.

Figure 1 shows a schematic drawing of a conventional image orthicon. Light from the scene is imaged onto the photocathode, where it releases a number of photoelectrons which is

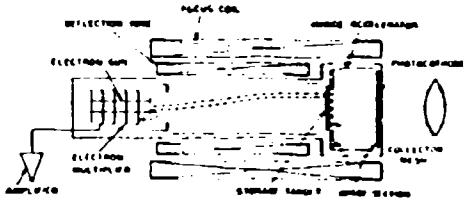


IMAGE ORTHICON
FIG. 1

proportional to the intensity of the light. The factor of proportionality is the photocathode sensitivity. These photoelectrons are imaged by a combination of an electrostatic field and axial magnetic field and strike the insulating target at high velocity. Since the secondary emission ratio is greater than unity, a positive charge image is developed at the target. The conducting collector mesh is fixed at a potential a few volts positive relative to the gun cathode. There is therefore a maximum signal which can be developed for a given target-mesh capacitance. Secondary electrons which originate from a target element which is at a higher potential than the mesh will not be collected, but will instead be redistributed to the target.

The electron beam, which is made to scan the target by a magnetic deflection, performs two functions in depositing negative charge at each target element to neutralize the positive charge produced by the photoelectron action. These charges should recombine within a frame time. The beam establishes an equilibrium potential at the target, in preparation for the integration of the next frame, and it reads out the signal stored at the target. The signal is extracted from the negative modulation on the return beam.

In approaching the problem of resolution in a camera tube, one begins by understanding that the tube is a system which is composed of cascaded imaging processes. Each of these processes is a low-pass filter with its own response characteristics. The response of the system is the response of these filters in cascade. The problem is attacked by studying and developing each of these processes.

In the image orthicon, there are three fundamental imaging or aperture processes: The image section, the storage target and the electron beam. In consideration of other camera tubes, other combinations would be considered.

The image of a point in the photocathode is confused into a disc at the target because of the photoelectron velocity distribution on emission. Electrons of any radial velocity execute helical trajectories about the magnetic field in an orbit time which depends only on the magnitude of the field, although the orbit diameters are proportional to the radial velocity. If there were no variation in axial velocity, all electrons originating in a point at the photocathode could be imaged to a point at the target, by adjusting the electron transit time to equal its orbit time. One actually focuses by making this match for an intermediate axial velocity; faster photoelectrons will not have reached focus at the target while slower electrons will have already passed through focus.

The formation of the beam spot is basically dependent on the same kind of imaging in a long magnetic field. However, the beam electrons approach the target at low velocity in the image orthicon. The lateral field presented by the charge pattern at the target is able to affect the beam image by deflection, and space charge can come into play, as has been observed in some of our experiments.

The aperture property of the target is determined by its lateral resistance, and is a measure of how well it can keep the charge pattern in place for a frame time.

*This research has been supported in part by The United States Air Force under Contract No. AF 33(616)-6682, monitored by the Electronic Technology Laboratory, WADD.

Another process, which is not fundamental to the image orthicon but which is important in the conventional operation of this tube, is the leakage of the magnetic deflection field into the image section. It perturbs the photoelectron trajectories and smears the image to the target. This effect is amenable to the same kind of response analysis as the other electron optical processes.

Before describing some of the methods and results in studying camera tube response, the methods of specification of resolution should be reviewed, following Schade.[1] The sine-wave response, $\mathfrak{R}\bar{\Psi}(N)$, of an imaging system is the relative amplitude of the output of the system as a function of spatial frequency, when the input is a bar pattern of sinusoidally varying amplitude and varying spatial frequency. The distance unit one generally considers in television is the picture height. The frequency unit is the line-number, N, which is the total number of bars in the pattern, both black and white, which can be fitted into the distance unit.

Responses to square-wave inputs are readily converted into sine-response. The sine-wave response is a very complete expression of the aperture response of an imaging process. The sine-wave response of the system at a given line-number is the product of the responses of its cascaded components at that line-number. An excellent single-number representation of resolution capability is an informational power integral across the sine-wave response:

$$N_e = \int_0^\infty \left[\mathfrak{R}\bar{\Psi}(N)\right]^2 dN.$$

Schade has demonstrated that imaging systems with differing contours in sine-wave response but with equivalent N_e's provide subjectively equivalent sharpness. The N_e for a system is closely approximated by the inverse quadrature of the N_e's of its components.

The N_e for a good average three-inch image orthicon operating near the knee in its transfer characteristic so that the effects of redistribution are not too strong is about 250 lines, or 280 lines per inch.

The usual specification of resolution in terms of limiting line-number is of little interest for these purposes, because of its ambiguity. Processes with the same limiting response can differ considerably in their response at intermediate line-numbers. Furthermore, one is concerned here with geometrical properties of imaging systems, whereas the observable limit in resolution at low signal levels is determined by the signal-to-noise ratio.

1. O. H. Schade, "Image Gradation, Graininess, and Sharpness in Television and Motion-Picture Systems," Part IV, Jour. SMPTE, 64, p. 593 (November 1955).

Figure 2 shows the basic experiment for determining the resolution capability of the image section process. The front end of the

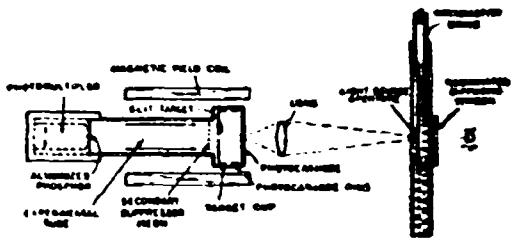


IMAGE SECTION APERTURE RESPONSE ANALYSIS
FIG. 2

tube is the image section of a three-inch image orthicon in which the collector mesh has been omitted and the insulating target has been replaced by metal foil which contains a slit a few ten-thousandths of an inch wide. A disc of light of a few ten-thousandths of an inch in diameter is imaged onto the photocathode and tracked across it with a micrometer drive at the source. The photoelectron image is in this way tracked across the analyzing slit. A mesh suppresses secondary electrons which originate in the target foil, and the photoelectrons which have passed through the slit and the mesh are accelerated and strike a phosphor layer backed with opaque aluminum. The light output from the phosphor is read with a photomultiplier. A line-transmission curve is traced out in this way. The Fourier transform of this function is the sine-wave response of the image section process in cascade with the responses of the analyzing slit and the light spot. The response of the slit is known from its width and the response of the light spot is determined by an independent measurement, so the sine-wave response of the image-section process can be extracted.

Figure 3 shows the results of such experiments on focus, compressed into N_e. The photocathode is the standard Ag-Bi-Cs_2-O surface, and a tungsten lamp at high temperature is the light source. The image section is fundamentally suited to very high resolution applications. N_e is about 1300 lines per inch with -500 or -700 volts at the photocathode in single-loop operation. The response decreases with lower photocathode voltage or photoelectron transit time, as is known. Although a simple theory predicts that the resolution is independent of

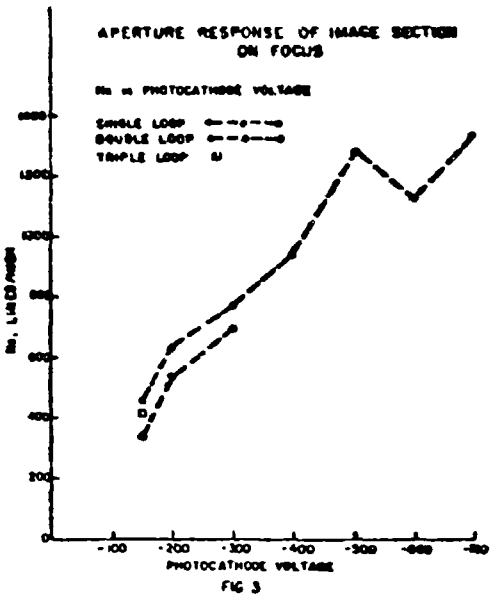


FIG 3

the number of cyclotron periods executed by the photoelectrons for a fixed transit time, it has been found that response is degraded by multiple loop operation.

These measured responses are three or four times higher than what will be estimated from a computation of the image diameter at the target from a point source in the photocathode with a range of emission velocity of one volt, if one assumes that the electron flux in the image is fairly uniformly distributed. The flux density is actually crowded sharply to the center of the point-image.

These results do not imply that the aperture response of an operating image section in an image orthicon always has this quality. Figure 4 shows the variation in image section response with deviation in the magnetic field from the focus condition. A one percent deviation in focus coil current can decrease the response by one-half. The focus field supplies must obviously be well-regulated. Furthermore, in order to obtain uniformly high response over the entire picture area, the orbit time for photoelectrons must be matched to the transit time within the narrow limits specified by these measurements for any point of origin in the photocathode. The curvature of the image field has been measured in a number of models of the three-inch image orthicon, and has been found to depend on the ratio of the image accelerator voltage to the photocathode voltage. When this ratio is too

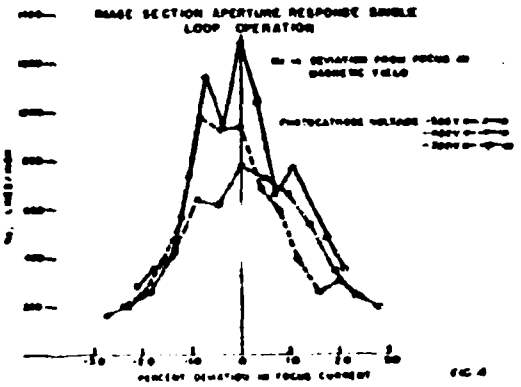
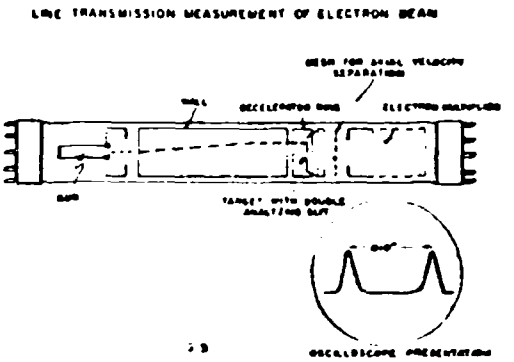


FIG 4

high, the image field is concave toward the photocathode, because the orbit time off-axis is relatively too short. When the ratio is too low, the image field is curved in the opposite sense. In an image orthicon which does not contain a field mesh, this ratio is assigned to minimize image distortion, and is too high for a flat image field. It has been found possible to produce a flat image field, coincident with the plane of the target simultaneously with low distortion by shunting off part of the magnetic field with a permeable ring around the image section. The deviation from focus out to the corners of raster corresponds to less than one-half percent of the focus coil current.

Experimental analysis of electron beams for camera tubes has been performed in two sections. Figure 5 shows one method, which is analogous to the image section experiment. The beam is

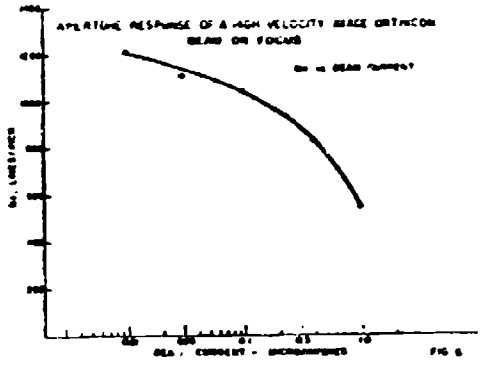


made to scan a thin metal target across two parallel narrow slits, one-hundreth of an inch

apart. Electrons which pass through the slits approach a fine mesh which has been used for axial velocity analysis. Electrons which can pass the mesh proceed to an electron multiplier. The presence of two slits provides a distance calibration for the analysis of the line-transmission function which is derived as a signal from this tube. Aperture response and the radial distribution of electron flux in the beam spot are determined from the line-transmission measurements. The structure of the beam is studied in this way in the absence of the lateral fields produced in the neighborhood of a camera tube target by the charge pattern it carries.

The second experimental system is a monoscope. A pattern of insulator in bar patterns of increasing frequency is formed on a metallic base. Aperture response of the beam may be measured from the television signal it produces when it scans the target. Targets may be made to operate at either high or low beam velocity. The signal is produced at high velocity by the difference in secondary emission ratios of the insulator and metal. At low velocity, the beam sees a difference in potential from the insulator and the target, and beam-bending effects may be observed.

Figure 6 shows the aperture response of the beam on focus at high velocity from an image orthicon gun. The data have been compressed



into N_e and are shown as a function of beam current. In the range of current required to operate an image orthicon at high signal level, the N_e is about 1100 lines per inch. This is significantly better than would be obtained from the imaging of the final limiting aperture in the gun, and the direct inference is that an electron cross-over within the gun smaller than the limiting aperture is imaged.

Resolution is poorer at low velocity. A decrease in resolution which has a behavior which can be attributed to space charge has been observed. An effect called "self-sharpening" which has been postulated for the low-velocity beam, according to which beam response will improve as the target voltage is decreased because low axial velocity may be correlated with larger distance from the beam center, has been observed only in relatively high current beams of about one microampere. The effect even here is small, and it could not be observed with smaller beam currents such as would be used in the image orthicon.

Figure 7 shows the reduction in response for the high velocity beam with variation in the electrostatic focusing field. The response of

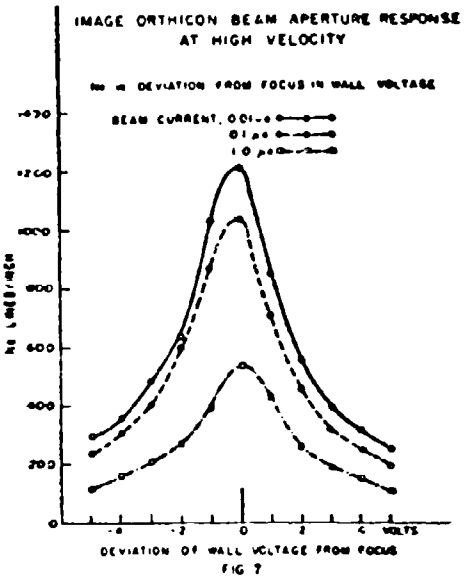


FIG 7

the electron beam from an image orthicon gun is much more sensitive than the image section process to variation in the focusing parameters. About one-half the response is lost with a one percent deviation in the electrostatic focusing voltage or with a one-half percent variation in the magnetic field.

These measurements present a major problem in the development of a high resolution television camera tube with a large format: the provision of an electron beam with high aperture response on focus and large depth of focus. The main effort in electron optics for this project is directed this way.

This work is being supported by The Wright Air Development Division, monitored by The Electronic Technology Laboratory, as part of a program being executed in collaboration with the RCA Electron Tube Division for the development

of an advanced television camera tube. This tube is proposed to deliver a combination of resolution capability and signal-to-noise ratio which is far beyond the generally recognised state of the art.

EXTENDING THE DYNAMIC RANGE OF CAMERA TUBES EMPLOYING RETURN BEAM MODULATION*

By: Messrs. A. Danforth Cope & Harold Borkan
RCA Laboratories, Princeton, N. J.

INTRODUCTION

Television camera tubes of varying complexity have been developed to convert the count of incident photons to an associated count of electrical charges accumulated at a storage target. The variation in the charge flux of each capacitive target element is sensed as potential variation when scanned by a fixed amplitude electron beam. As every storage element is sampled, the charge exchanged between beam and target not only wipes out the stored information, but generates the video signal. In iconoscope, orthicon, or vidicon tubes the photosensitive element is also the storage element. When amplification of the photon flux is desirable, single stage image sections or multi-stage intensifiers have been employed between photo-emitter and target.

With a particular incident photon flux, the signal-to-noise ratio achievable by an ideal (noiseless) device is dependent only on the quantum efficiency of the photoelement. All practical devices fall short of the ideal because they introduce noise in storing an amplified signal at the target, and in converting this stored information to a video signal. This paper will examine the operation of the image orthicon to determine the limitations attributeable to each noise source. The extent of improvement afforded by an alternative scanning method employed in the Image Isocon will also be examined and experimental comparisons reported.

IMAGE ORTHICON AND IMAGE ISOCON CAMERA TUBES

Among commercially available camera tubes, the image orthicon (Fig. 1) is operable at the lowest scene illuminations. In this device the electrons emitted by the photocathode are multiplied by the secondary emission of the bombarded face of the target. For low levels of illumination the ratio of target charges to incident photons, which might someday be as large as 5, is typically 0.6 or 0.7. A factor of 1.2 has been achieved in recent improved tubes.

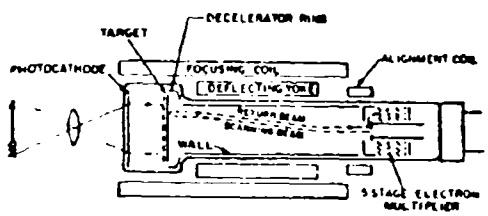


FIG. 1 IMAGE ORTHICON

The orthicon scanning beam is collimated to contain a minimum of radial velocity components and is directed along an axial magnetic field of the order of 75 gauss. The axial velocity is reduced to essentially zero as the electrons approach the target so that sufficient negative charge is deposited at each target element to neutralize the positive charge which has accumulated since the last scan. The reflected beam current, now modulated according to the number of charges subtracted at the target, is directed into a multi-stage electron multiplier which provides sufficient amplification for the noise in the output signal to exceed the noise of the video amplifier. The spread in axial electron velocities within the beam limits the maximum signal modulation to less than 30% in most tubes.

Signal generation by scattered electron beam modulation which is unique to Isocon scan is quite different.[1] Not all the electrons striking the target with axial velocities of only a few volts will stick. Some are scattered non-specularly and mix with the specularly reflected electrons. These scattered electrons have a broad spectrum of transverse velocity components whereas the reflected electrons have a restricted spectrum. In the axial magnetic field each electron will, at the cyclotron frequency, traverse a helical path whose radius is proportional to the square root of the radial velocity. Because of this, the two classes of electrons with a common point of origin are separable. At the plane where the maximum spatial separation occurs an apertured electrode is placed so as to intercept the reflected electrons while allowing the majority of the scattered electrons to enter the multiplier. With proper electrode adjustment, the output current in the dark will be only a small fraction of the highlight output current.

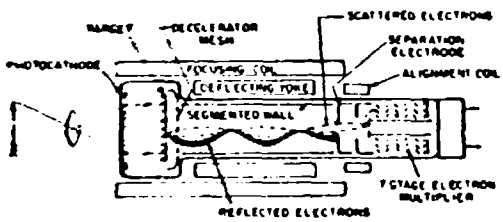


FIG. 2 DUAL SCANNING MODE CAMERA TUBE - ISOCON SCAN

Dual operation tubes, Fig. 2, have been used for comparing orthicon and Isocon scan of identical target charge patterns. The wall electrode has been divided into four quadrants to which independent potentials can be applied. The mesh on the scanned side of the target provides the

*This work was supported by the United States Air Force under contract Nos. AF33(616)-5728 and AF33(616)-6497, monitored by the Electronic Technology Laboratory, Wright Air Development Div.

[1] The Image Isocon, P.K. Weimer, RCA Review, Vol. XI, No. 3, pp. 366-386, (Sept. 1949).

concentrated, uniform decelerating field at the target which is required to give a return beam with minimum scan. When operated with an axially aligned beam and the same potential on all wall quadrants, the scan is essentially that of the normal image orthicon.

Isocon scan is achieved by setting up a transverse electric field in the region of the wall quadrants. The resultant radial velocity of the beam is such that the reflected electrons are intercepted by the separation electrode. The steering of the beam can be assisted by the transverse magnetic field produced by the alignment coil. By careful adjustment a uniform signal from the entire target area may be obtained from the scattered electrons which, at the plane of the separation edge, lie closer to the tube axis than the reflected electrons. With this tube design signal modulation in excess of 90% has been obtained.

SIGNAL-TO-NOISE ANALYSIS

The range of light levels in a scene that is usefully transduced by a television camera tube is limited because of the noise in the low-light regions. The following simplified analysis yields the signal-to-noise ratio as a function of light level for the image orthicon[2] and the Image Isocon. In addition, tube design data pertaining to signal-to-noise ratio of these tubes is derived.

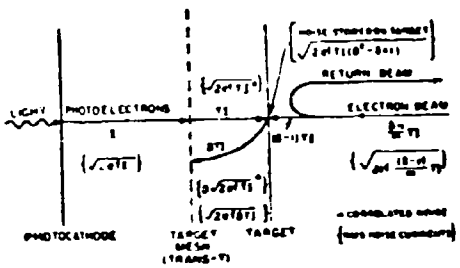


FIG 3 IMAGE ORTHICON CURRENT FLOW

The image orthicon and the Image Isocon have similar image sections and will, therefore, for the same incident illumination have stored on the target the same signals accompanied by the same noise. However, the noise contributed by the scanning process will differ in the two tubes. Figure 3 shows the signal currents and RMS noise currents in the image orthicon. It is assumed that the current generated at the photocathode is represented by I, the target mesh transmission by T, and the secondary emission ratio for photoelectrons at the target by δ. This secondary emission process produces a stored signal on the target of $T(\delta-1)I$ and an RMS noise stored on the target of $\sqrt{2ef\,TI\,(\delta^2-\delta+1)}$, where e is the electronic charge in coulombs, and f is the bandwidth in cycles per second.

If it is assumed that the beam current landing on the target is correlated to the stored charge, the scanning process will not degrade the signal-to-noise ratio, but the total shot noise contained in the incident beam will appear along with the output signal. The beam modulation, m, is defined as the ratio of the current which lands on the target to the incident beam current. Therefore, in these terms the RMS shot noise current contained in the incident beam will be $\sqrt{2ef\frac{\delta^2}{m}TI}$. The signal-to-noise ratio at the entrance to the electron multiplier is equal to the signal extracted from the target divided by the RMS noise current which is composed of the noise stored on the target and the shot noise in the beam. Therefore,

$$\frac{S}{N_{IO}} = \sqrt{\frac{I}{2ef}}\sqrt{\frac{T(\delta-1)}{\delta\frac{1}{\delta-1}+\frac{1}{m}}} = F_1(I)\cdot F_2(T,\delta,m).$$

Note that F_1 is the signal-to-noise ratio of the photoelectrons and that F_2 contains only parameters which pertain to storing the information and deriving the video signal. In other words, F_1 gives the maximum possible signal-to-noise ratio available for a particular value of photocurrent, while F_2 determines the fraction of F_1 that is realizable at the multiplier input. It is easily shown that degradation within the electron multiplier is negligible, generally less than 10%.

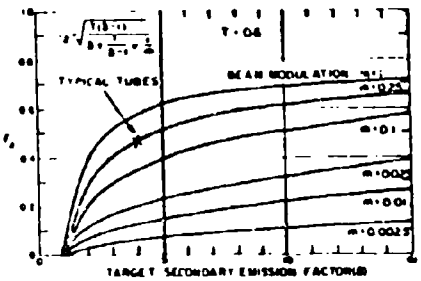


FIG 4 IMAGE ORTHICON SIGNAL-TO-NOISE MODIFYING FACTOR (F_2)

The image orthicon signal-to-noise modifying factor, F_2, is plotted in Fig. 4 as a function of the target secondary emission factor and beam modulation. Standard image orthicons with $\delta = 4$ and $m = 0.25$ produce a maximum F_2 of about 0.46. Increasing beam modulation to unity or δ to 7.5

[2] A more complete signal-to-noise analysis of the image orthicon has been made by E.G. Ramberg, "A Theoretical Analysis of the Operation of Flying Spot & Camera Tube Microscopes in the Ultraviolet," IRE Transactions on Medical Electronics, Vol. PGME-12, Dec. 1958.

would improve the signal-to-noise ratio about 25%.

Figure 3 shows the currents flowing within an Image Isocon. While the signal and noise stored on the target is the same as in the image orthicon, the Isocon generally realizes an additional signal gain since the number of scattered electrons exceeds the number which remain on the target to neutralize the stored charge. Unlike most image orthicons, the Isocon requires a decelerating mesh on the beam side of the target. The scatter gain is represented by α and the mesh transmission by T_d. The signal collected by the electron multiplier is αT_d times the signal stored on the target. Calculations show that the signal-to-noise degradation due to the decelerator mesh is only a few percent and is therefore neglected.

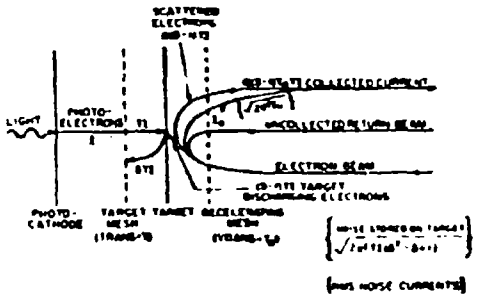


FIG 5 IMAGE ISOCON CURRENT FLOW

Along with the collected signal a constant fraction of the beam current represented by I_o, is present. The shot noise of this current together with the noise contribution from the target are used to determine the signal-to-noise ratio at the multiplier entrance, yielding:

$$\frac{S}{N}_{II} = \sqrt{\frac{I}{2ef}}\sqrt{\frac{T(\bar{\alpha}-1)}{\bar{\alpha}+\frac{1}{\bar{\alpha}-1}+\frac{1-\mu}{\mu\alpha T_d}}} = F_1(I)\cdot F_2'(\bar{\alpha},\mu,T_d).$$

where μ, the signal modulation, is the ratio of peak signal current to the total collected current. As with the image orthicon, F_1 is the signal-to-noise of the photoelectrons. The F_2' dependence of the Isocon signal-to-noise modifying factor on α, μ, and $\bar{\alpha}$, for usual values of T and T_d is plotted in Fig. 6.

DYNAMIC RANGE

The Image orthicon and Image Isocon signal-to-noise expressions have been used for comparison of dynamic range performance. Figure 7 shows, as a function of light level, the relative signal-to-noise performance of typical tubes and ideal tubes with perfect highlight modulation. The graphs apply to operation below the knee of the light-transfer characteristic. It has been assumed that the camera tube adjustments are

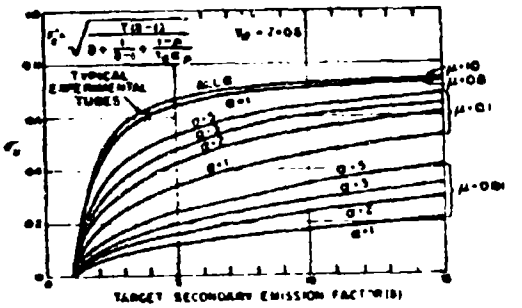


FIG 6 IMAGE ISOCON SIGNAL-TO-NOISE MODIFYING FACTOR (F_2')

optimumized for the picture highlights and no further adjustments made. Note that the dynamic range expected from a realizable Image Isocon is superior to that of the ideal image orthicon and is about 8 times better than existing tubes.

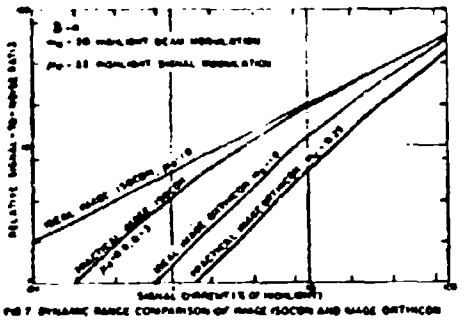


FIG 7 DYNAMIC RANGE COMPARISON OF IMAGE ISOCON AND IMAGE ORTHICON

Shown in Fig. 8 is the measured signal-to-noise performance of an experimental dual-mode tube having a 2.5 mil spacing between collector mesh and target. The measured signal-to-noise ratio was taken as peak-to-peak signal divided by one-sixth of the peak-to-peak noise. The target secondary emission was approximately 5, the highlight orthicon beam modulation 13%, the highlight Isocon signal modulation was 90%, and the Isocon scatter gain about 2.5. Two sets of data are shown for both orthicon and Isocon scanning of the same target information. In each mode the lower curve indicates the dynamic range, since a fixed beam current adequate to discharge the highlight was maintained.

The upper curves show data taken with the beam current adjusted at each point for highest modulation consistent with adequate discharge of the stored target charge. This experimental data shows a factor of 10 greater dynamic range for Isocon scan, and a 3 to 5 times improvement in

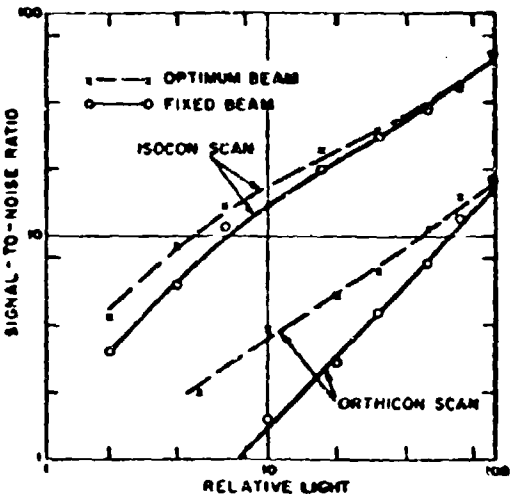


FIG. 8 EXPERIMENTAL SIGNAL-TO-NOISE PERFORMANCE

measured signal-to-noise ratio with optimum beam setting. The experimental data agrees well with the analysis when measured parameters are used in the formulas.

APERTURE RESPONSE

Limiting resolution of an imaging device such as a camera tube serves only to indicate the upper frequency of line pattern which can be detected. Experience has shown that the better of two transducers with the same limiting response is that one which shows the less rapid drop-off in signal response with increased pattern frequency.

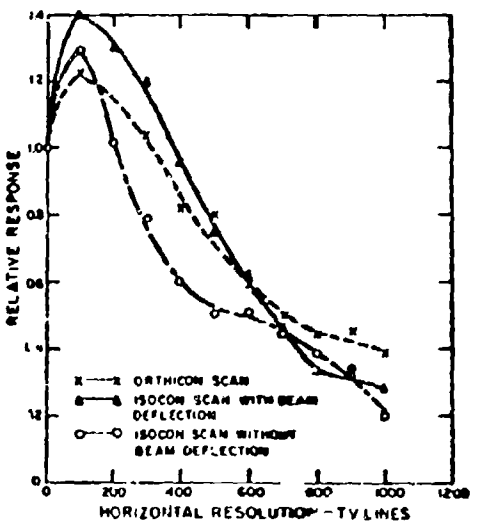


FIG. 9 SINE-WAVE APERTURE RESPONSE

Figure 9 gives the sine-wave response as a function of horizontal resolution for orthicon and Isocon scan in a dual mode tube which is illuminated to full charge storage. The scan was restricted to the center of the target to minimize scanning modulation. When beam deflection effects at the target are permitted to give a differentiated signal along a given direction (in this case along a horizontal line), the response can be enhanced. The response, for orthicon scan is bracketed by the response for normal and for differentiated Isocon response. At low light levels, Isocon operation yields better resolution because of the superior signal-to-noise ratio.

QUANTITATIVE EVALUATION OF DETECTION

A quantitative measure of the ability of any optical transducer to extract information from a scene has been established by A. Rose.[3] A figure of merit dependent on the brightness, contrast and size of test objects has been employed in Fig. 10 to relate tube performance to the illumination on the tube face. With a given tube, a one digit change in the figure of merit results when the photon flux is changed by a factor of four. With larger signals the rate of change of the figure of merit for either mode of operation is that established by the photon flux. At the low light level where the figure of merit drops rapidly the performance limitations within the tube are dominant and the device is approaching its detection limit. The high beam modulation performance of Isocon operation accounts for its superior results.

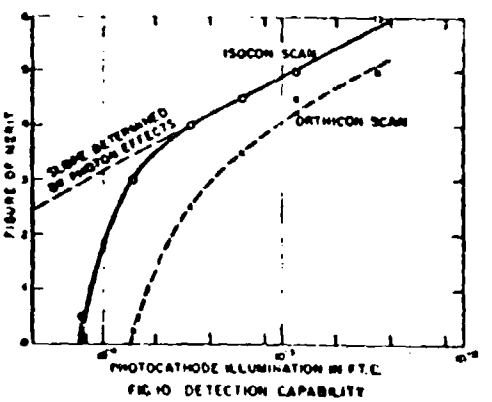


FIG. 10 DETECTION CAPABILITY

[3] A. Rose, Television Pickup Tubes and the Problem of Vision, Advances in Electronics, Vol. 1, 1948.

CONCLUSIONS

Significant advantages generally obtained from the Isocon scan as compared with orthicon scan are:

1. A Factor of 2-3 improvement in signal-to-noise ratios over a broad range of light levels below the knee.
2. At least an order of magnitude reduction in the low light level threshold.
3. An order of magnitude greater dynamic range.
4. Comparable high light resolution and superior low light resolution.

For low light level operation with a wide range of brightness in the scene the increased operating complexity of scattered electron beam modulation is more than offset by the superior results obtained.

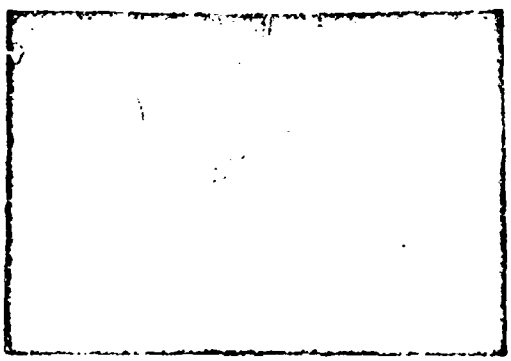
IMAGE ORTHICON TUBES AS IMAGE INTENSIFIERS

By: Mr. Nils Svenson and Mr. James Parton, Warfare Vision Branch, USAERDL, Ft. Belvoir, Va.

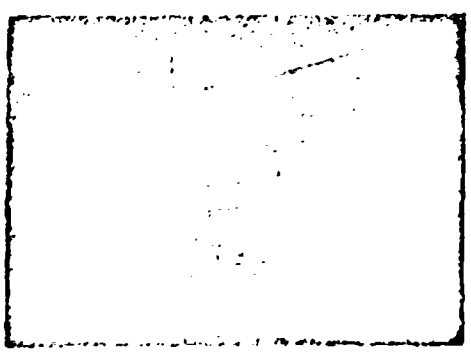
I. Introduction

During the past year the Warfare Vision Branch of USAERDL has been evaluating various types of image orthicon tubes as image intensifiers in low light level television systems. This is part of a research and development program to provide a night time remote viewing capability for the U.S. Army.

The standard broadcast studio image orthicon (the 5820) has good performance at high light levels associated with normal broadcasting techniques, but drops off badly at light levels occurring in nighttime scenes. Accordingly, the present evaluation has concerned itself primarily with two other image orthicon type tubes showing greater potential for nighttime viewing. One of these is an image orthicon using standard image orthicon geometry but incorporating a thin film magnesium oxide target in place of the conventional glass target (see fig. 1). This special



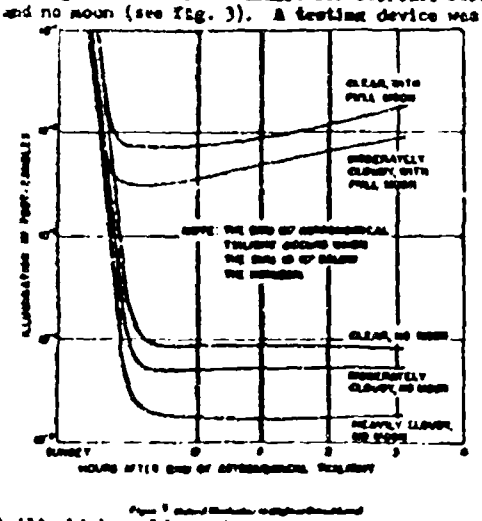
target provides increased sensitivity at low light levels due to a combination of increased secondary emission ratio and high lateral resistivity (but adequate through-conductivity along grain boundaries). The other image orthicon is a tube using an additional image section in the front end acting as a preamplifier for the conventional image section (see fig. 2). This additional image section is essentially an electrostatically focused image converter tube coupled to the conventional image orthicon through a thin glass membrane with a phosphor on one side and a photocathode on the other. The glass membrane is flat and is situated with regard to the glass target exactly where the front faceplate is in a conventional image orthicon. The scene being viewed is imaged on the first photocathode by a lens. Electrons are emitted from the photocathode in a manner similar to the image orthicon, but these are electrostatically focused on the thin glass membrane. The electrons are accelerated through 15-20 KV before striking the phosphor, which in turn emits light causing photoemission from the second photocathode. This stage acts as an intensifier since more electrons are emitted from the photocathode than originally struck the phosphor.



Both types of tubes use a wide spacing of target to mesh to reduce target-mesh capacitance and provide a greater signal voltage for a given amount of charge on the target at low light levels

II Experimental Arrangement

Scene illuminations in typical nighttime conditions vary from 10^{-2} foot-candles for full moonlight to 10^{-5} foot-candles for overcast skies and no moon (see fig. 3). A testing device was



built which could provide controlled and carefully measured highlight brightnesses through this range. (Scene brightness equals scene illumination times the scene reflectance). The tubes were evaluated using highlight photocathode illumination as the independent variable, rather than highlight scene brightness, to avoid incorporating the lens used into the performance characteristics of the tubes. By geometrical optics one can show that the photocathode illumination E in foot-candles is related to the scene brightness B in foot lamberts

$$E = \frac{B}{4f^2} \frac{\tau_L \tau_F}{(1+m)^2} \quad \text{(for small angles of view)}$$

where f = f/number of lens
m = magnification of image
τ_L = transmission of lens
τ_F = transmission of neutral density filters used to attenuate the light.

The television camera itself is a standard RCA TK-31A field camera system modified by using a low noise video preamplifier. This preamplifier has a cascode input circuit with a Western Electric 417A triode in the first stage. The preamplifier noise is equivalent to 3×10^{-9} amperes r.m.s. noise current through the 18 Kohm load resistor of the image orthicon when the preamplifier is operating in the camera but with no beam current flowing from the image orthicon.

A refrigerating unit with a blower is used to regulate the target temperature of the image orthicon tubes. A thermistor in contact with the target pin socket of the image orthicon camera acts as the temperature sensing element. The target temperature is varied in order to change the target resistivity to determine the effects on picture resolution, time lag, etc., at low light levels. The heat conductive path to the target through the target pin is poor, and since the target is heated by radiation from the tube filaments the exact target temperature is unknown. However a resettable value of target pin temperature can be maintained for purposes of comparison.

The image orthicon using the magnesium oxide target requires no other auxiliary equipment for operation. Since the image orthicon with the intensifier section is longer than the conventional tube, a mechanical modification of the camera as well as a high voltage supply for the intensifier section is required.

VI Results

Since the low light level performance of these tubes is of primary interest, measurements were carried out in the range from 10^{-4} to 10^{-7} foot-candles photocathode highlight illumination. Quantities measured were limiting resolution, signal to noise ratio, and signal current output for 20°C and 35°C bulb temperatures. Aperture response curves were taken at several light levels at the two temperatures. All these measurements were made using light of 2870°K color temperature. A conventional 525 line picture, 30 frames and 60 fields per second, was used in taking the data. The bandpass of the video amplifier is 10 mc.

The image intensifier orthicons are divided into three classes. The first of these is the class of tubes with both photocathodes multialkali (S-20), requiring the use of a special target glass compatible with the alkali elements in the photocathode. These are specified as "Type A" tubes. The second class, the "Type B" tubes, contain a multialkali first photocathode, an S-11 second photocathode, a standard target glass, and a low power heater for the thermionic cathode to decrease thermal radiation to the target. The third class, "Type C" tubes, are the latest tubes and have a glass target with higher resistivity but are otherwise identical with the "Type B" tubes. All tubes use a P-11 phosphor on the intensifier membrane. At least three tubes of each type were evaluated.

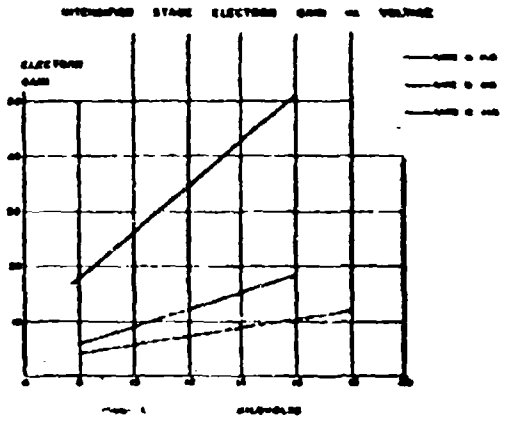
Only one of the magnesium oxide target image orthicons was evaluated, to compare the performance with the image intensifier orthicon. However, the test results indicated this tube was a "very good" tube when compared with the manufacturer's curves. This tube uses an S-10 photosurface, with a typical response of 50 to 70 μamperes per lumen. The important differences in these tubes are tabulated below.

Tube Type	First Photocathode	Second Photocathode	Target Glass	Low Power Heater
Type A I.I.O.	S-20	S-20	multialkali compatible	No
Type B I.I.O.	S-20	S-11	standard	Yes
Type C I.I.O.	S-20	S-11	higher resistivity	Yes
MgO Target I.O.	S-10	---	---	No

The front photocathode sensitivity of the image intensifier orthicons ranged between 110 and 150 μamperes per lumen (using 2870°K light). The gain of the intensifier section as a function of accelerating voltage is shown in figure 4. Electron gain is defined as the current emitted from the 2nd photocathode divided by the current striking the phosphor. It is highest for the tubes with two multialkali photocathodes by a considerable margin. The maximum possible gain at 16 KV for the type A tubes is about 90. About 50 has been achieved to date. For the type B & C tubes (using a P-11 phosphor and an S-11 2nd photocathode) the maximum gain would be about 70. Here the typical value shown (for type C) is 18, although one tube was as high as 23.

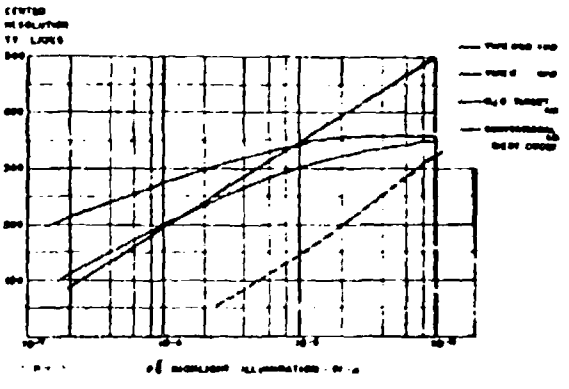
The limiting resolution measurements resulted in the curves shown in figure 5. The resolution was measured in the central area of the tubes. The electrostatic image section of the I.I.O. has a characteristic drop off in resolution toward the edge so that the 200 line vertical bars at the edges of the Retma resolution pattern could not be resolved even at the highest light levels used in the tests.

The magnesium oxide target tube and the types A & B image intensifier orthicons, for resolutions of 250 lines or less, produce the same



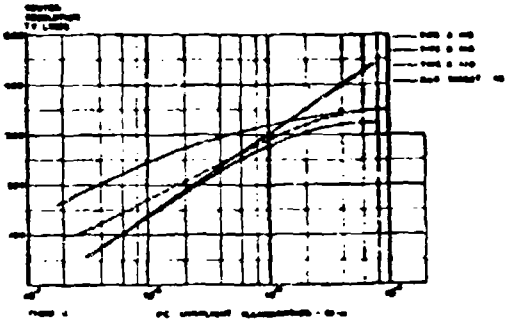
resolution picture at 1/20 - 1/30 the light level required by a good conventional image orthicon. The type C tube requires about 1/100 of the light necessary for the conventional image orthicon in this region. This is close to what one would expect from a combination of multialkali photocathode and intensifier section. A three times gain in photocathode sensitivity over the conventional I.O. and a maximum gain of 70 in the intensifier membrane would provide a total gain of about 200 over the conventional image orthicon. Further improvement could come in new target developments or by increasing the scanning beam modulation.

The leveling off of resolution at the higher light levels is due to the construction of the intensifier section. The intensifier membrane is glass .001" or less thick with the phosphor coating on one side and the photocathode on the other. The light from a point image on the phosphor undergoes some spreading in passing through the glass and is seen as a disc on the photocathode. Scattering of light in the phosphor, and reflection of light from the photosurface back to the phosphor and its rescatter to the photocathode limit the maximum resolution. The coupling of the magnetic field of the conventional image section into the electrostatically focused portion of the tube, as well as possible distortion of this focusing field by the large pieces of Kovar in the tube may have a defocusing effect.



These resolution measurements shown are typical values. Variations exist from tube to tube, and in some cases from place to place in the central area of the same tube. However, all the measurements of a given tube type fall in a band surrounding the line shown as typical. The error in photocathode illumination is estimated at ±25%.

When these measurements were repeated at a bulb temperature of 35°C (fig. 6) the resolutions of all the tube types decreased to some degree.

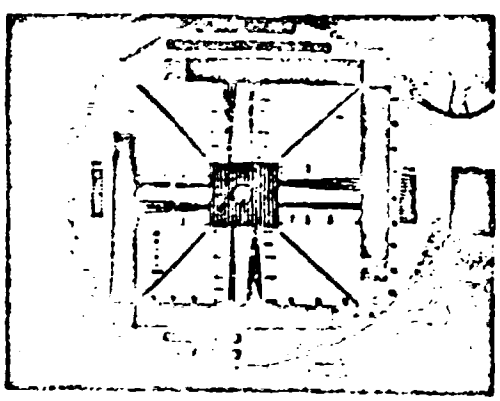


One notes that the resolutions of type A & B tubes are the same at 20°C, but that the type A tubes suffered a greater decrease at 35°C, especially at the lower light levels. This was the reason for changing to the S-11 2nd photocathode and the consequent higher resistivity glass target. The low power heater for the electron gun cathode introduced at the same time also played a role. As long as the type A tubes were kept cold, their performance was fairly good, but deteriorated upon warming up. These tubes lost about 60 TV lines resolution at 10^{-6} ft.-c. while the type B tubes lost about 30 lines resolution. The magnesium oxide target tube lost about 50 TV lines resolution over its whole range at 35°C as compared to 20°C. The latest, or type C, tubes lost about 25 TV lines. The high light level resolution was not as severely affected by changing the bulb temperature; for most I.I.O. tubes it didn't seem to make any difference. Here it must have been the intensifier section of the tube rather than the target resistivity which was limiting the resolution.

It should also be noted that these are "static" measurements; they do not involve moving objects. Hence effects such as time lag and stickiness do not appear. It may be said, speaking qualitatively, that the magnesium oxide target tube is considerably more "lazy" at low light levels than the image intensifier orthicon. Even the type C tubes, with the highest resistivity glass targets are not nearly as bad in this respect. This was borne out in a recent field demonstration when a scene illuminated only by starlight (10^{-4} ft.-c. measured scene illumination) was viewed through an f/1.85 lens using both the magnesium oxide target tube and a type C I.I.O. Moving objects in the field of view were reproduced very well by the I.I.O., while the magnesium oxide target tube showed a trail behind the object,

like a blur, and a rapid movement would produce either a ghost-like image or no image at all.

Photographs of the kinescope presentation of two of the tubes at two different light levels were taken to illustrate the picture quality. Figure 7 shows the picture produced by the magnesium oxide target image orthicon at 1.6×10^{-4} ft.-c. photocathode highlight illumination. Figure 8 was made with the same tube at 2.5×10^{-6} ft.-c. photocathode highlight illumination. Figure 9 shows the picture produced by a type C image intensifier orthicon at 1.6×10^{-4} ft.-c. photocathode illumination. The overall uniformity and clarity of the MgO target tube picture at this light level is strikingly apparent when compared to the type C I.I.O. Figure 10 shows the picture produced by the I.I.O. at 2.5×10^{-6} ft.-c. photocathode illumination. Here very little degradation of the picture has taken place despite the large decrease in light level. These photographs were taken at 1/5 second exposure to simulate the integration time of the eye. The bulb temperature for both tubes was 20°C. The vertical lines in each photograph and which are most apparent in figure 8 are due to pickup in the camera, and have nothing to do with the tubes.

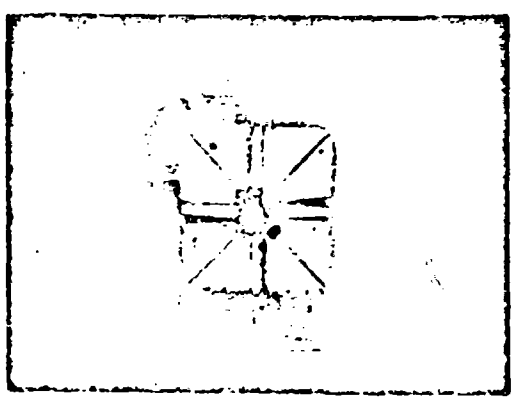


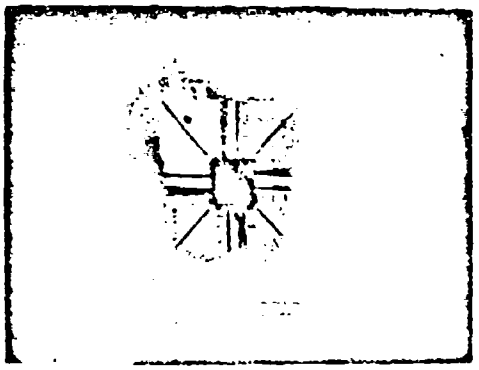


A Tektronix type 535 oscilloscope connected as a line selector oscilloscope was used to measure the signal to noise ratio, signal current output and aperture response. For the signal and noise measurements, the line selector is used to give a trace across a large area black to white transition in the picture. The peak signal is the black to white signal amplitude with the noise averaged out. The peak-to-peak noise is measured in the black region. Then signal to noise ratio is defined as peak signal to r.m.s. noise where r.m.s. noise is taken as 1/6 peak-to-peak noise (fig. 11). The beam current was adjusted at each light level to just discharge the target, so that no unnecessary beam noise was introduced. Even at the very lowest light levels (i.e. about 10^{-7} ft.-c.), the noise in the beam current required to just discharge the target was 2 to 4 times the noise due to the load resistor and video preamplifier. The total noise, i_t, is due to two contributions, the beam noise, i_b, and the amplifier noise, i_a. Since both are random and not interdependent, they add as

$$i_t = \sqrt{i_a^2 + i_b^2}$$

The total noise is measured as described above and the amplifier noise alone is measured when the image orthicon scanning beam is cut off.

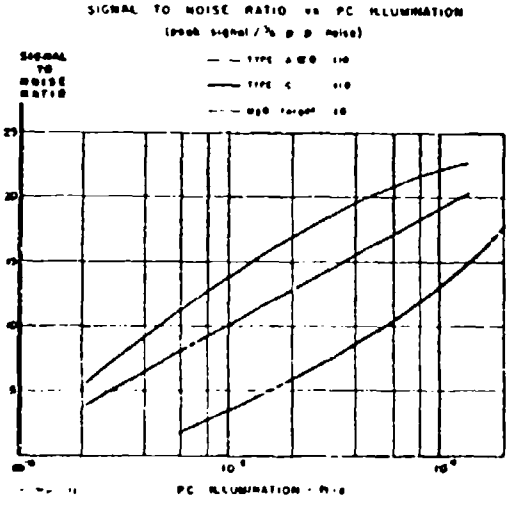




To determine the signal current output, the voltage read on the oscilloscope was divided by the gain of the preamplifier and by the image orthicon tube load resistor, in this case 13 Kohms. The gain of the preamplifier was measured at 20kc (fig. 12).

There was no evidence of temperature dependence of the signal current output or signal to noise ratio for any of the tubes tested (that is between 20°C and 35°C operation).

Aperture response curves were measured using a set of vertical bar patterns of various line numbers. These were square wave bar patterns, i. e. the brightness varied in a step function manner as one scanned across the chart from black bar to white bar. The results are summarized in the table below. At the highest light level at which tests were made, 1.7×10^{-4} ft.-c., a higher equivalent line number, N_e, (calculated by integrating the square of the relative response) was obtained for the tubes at 35°C, even though the limiting resolution was as good or better at 20°C. The type A tubes were not measured at this light level for fear of burning in the image on the target. At 2.0×10^{-6} ft.-c., only the type C tubes had a high enough signal to noise ratio to make possible an aperture response measurement.



Tube	Bulb Temp	Equivalent Line Number N_e, TV Lines, at each photocathode illumination, ft.-c. 1.7×10^{-4}	2.0×10^{-5}	2.0×10^{-6}
Type A I.I.O.	20°C	——	60	——
	35°C	——	48	——
Type B I.I.O.	20°C	92	60	——
	35°C	100	68	——
Type C I.I.O.	20°C	135	100	55
	35°C	150	96	50
MgO target I.O.	20°C	92	90	——
	35°C	125	75	——

IV Ideal Performance

Ideal performance for any type of picture pickup device is performance limited only by the finite number of available light quanta. Fluctuations in the number of photons received from the scene being viewed constitute a fundamental noise limit to the information available from the scene.

Rose[1] has derived a relationship for the ideal performance of any picture pickup device using as a test object a dark square on a lighter background. A threshold signal to noise ratio of 5 is considered necessary for an observer to just detect the presence of the square against the background, i.e. to resolve the square. Then Rose relates the background brightness B, the square to background contrast C ($C = \Delta B/B$), the diameter lens used by the pickup device , the angle subtended by the square at the lens, the exposure time, t, and the quantum efficiency of the primary photoprocess.

This relationship was converted at this Laboratory to a form more suitable for use with image orthicon tubes, using geometrical optics and assuming a 4:3 picture aspect ratio:

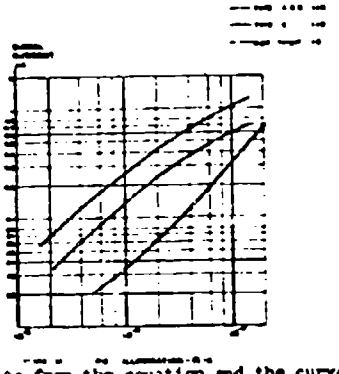
$$E = \frac{4k^2eR^2}{3C^2StA}$$

where E = photocathode illumination due to background, ft.-c.
k = threshold signal to noise ratio = 5
e = electron charge, coulombs

[1] Television Pickup Tubes and the Problem of Vision-A. Rose-Advances in Electronics Vol. I (Academic Press, New York, 1948).

N = limiting resolution, i.e. the number of squares which can just be detected against the background which will fit in the picture height.
C = contrast, $0 \leq C \leq 1$
S = primary photocathode sensitivity, amperes/lumen (white light)
t = exposure or integration time, seconds
A = primary photocathode picture area, $ft.^2$

Here lens characteristics do not appear, and we have the resolution of the tube related directly to the image contrast and photocathode illumination. This equation is plotted for different contrasts in figure 13, for 100 μamperes/lumen PC sensitivity and 1/5 second integration time (the value for the eye).



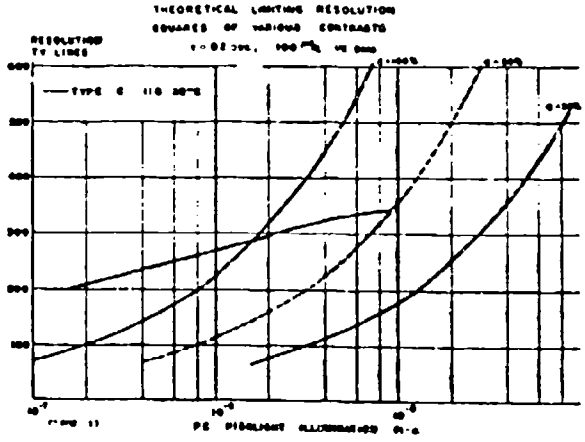
We note from the equation and the curves that for a given resolution, the photocathode illumination goes inversely as the square of the contrast. Typical night scenes may produce highlight photocathode illuminations as low as 10^{-6} ft.-c. even with a fast lens. Such scenes will have contrasts of 25-50%. Looking further at the curves we notice an apparent paradox. The type C image intensifier orthicon resolution at 20°C is plotted to the same scale, and exceeds the calculated theoretical resolution of an ideal tube in the region below about 10^{-6} ft.-c. PC illumination. Higher PC sensitivity, say 150 μamperes/lumen could account for only a small part of the difference. Probably the main reason for the discrepancy is the different way of measuring the tube resolution. In a TV Retma chart a pattern of converging alternate black and white lines is followed by the eye until the lines can no longer be resolved. This is called limiting resolution, and the eye no doubt gains some advantage in first, seeing several black and white lines adjacent to one another, and second, in being able to follow the pattern from where it is clearly visible. To get a real comparison one should actually try to distinguish a square against a uniformly illuminated background using the I.T.O. The theoretical calculation assumes that the eye is trying to detect the one information bit out of the background, and is not relying on adjacent geometrical shapes for clues.

One point which can be inferred from the curves is that the presently existing tubes are close to being photon noise limited (or more accurately photoelectron fluctuation noise limited, since quantum efficiencies are considerably less than one). One might guess that they are within a factor of ten of the ultimate in tube performance possible using existing photosurfaces of 10% quantum efficiency.

V Conclusion

This paper has summarized the results of a series of measurements on existing television camera tubes designed for low light level operation. Further advances in this field depend on the success of current research and development work now being pursued by various organizations. Among items under development is an image intensifier orthicon with a magnetically focused intensifier section. Concurrently with this a study of new type targets and improvements in the intensifier membrane are being carried on. This should provide a shorter tube than the present electrostatically focused version, with uniform image quality over the entire picture area, and improve the resolution at both high and low light levels. The first tubes are expected early next year. Attempts are being made to mount a fiber optic faceplate having 1 mil or smaller diameter fibers in an image orthicon. This will allow mechanical coupling of an image converter tube with a fiber optic output phosphor screen directly to the image orthicon, and should yield a large increase in sensitivity without an appreciable decrease in resolution. This replaces the intensifier membrane with two fiber optic coupling elements. Research in methods for reducing scanning beam noise, and a miniaturization program to provide smaller image intensifier orthicons are also in progress. The next twelve months should see considerable improvement in the state of the art of low light level television.

The authors would like to thank Myron W.Klein, John Johnson, and John Moody of the Warfare Vision Branch at USAERDL for their continued interest and for many helpful suggestions and discussions.



LIGHTWEIGHT INERTIAL SYSTEMS

By: Mr. Robert E. Marcille, Electronic Equipments Division, Litton Industries

Accurate lightweight inertial systems which are applicable to missiles and manned or unmanned aircraft are in quantity production at the Electronic Equipments Division of Litton Industries. The CEP of one mile per hour that is attained in several versions of these systems is comparable to the CEP achieved by systems that require much heavier equipment. Litton's lightweight systems now extend the use of inertial technique to applications where formerly the weight of inertial systems prevented or deterred their use. Litton lightweight inertial systems are currently being employed in the following U. S. Navy aircraft:

WF-2	—	Carrier-Based Airborne-Early-Warning Aircraft
P3V	—	Antisubmarine-Warfare Patrol Aircraft
W2F	—	Carrier-Based Airborne-Early-Warning-and-Control Aircraft
A2F	—	Carrier-Based Attack Aircraft

Foreign applications of these lightweight inertial systems (for NATO commitments) include their use in the F-104G and CF-104 Strike Fighters for the West German Ministry of Defense, and the Royal Canadian Air Force, respectively.

Litton's inertial guidance systems are also included in the configuration of missiles and drones designed by several weapon system contractors and under consideration by Department of Defense agencies.

Several hundred Litton systems are currently on order. Although these systems are built for different applications and therefore require different configurations, computers, displays and input-output mechanizations, all these systems are built around the production P-200 inertial platform. This 30-pound platform, shown in Figure 1, is 14-1/2 inches long and has a maximum diameter of 9-1/2 inches.

The platform has a four-gimbal, three-axis configuration that prevents gimbal lock, thereby permitting full maneuverability of the carrier vehicle. The stable element of the platform includes two two-degree-of-freedom, floated gyros, and three pendulous-type, torque-balance accelerometers. The two gyros are mounted on the stable element with their spin axes horizontal and at 90° to each other; the three accelerometers are mounted with their sensitive axes oriented to the horizontal X and Y axes and to the vertical Z axis of the stable element, respectively.

The stable element erects itself to the local vertical by nulling the outputs of the X and Y accelerometers, and aligns itself to true north by automatic gyrocompassing. The systems are normally mechanized so that the stable element is always maintained at local level and in the selected alignment, regardless of aircraft maneuvers.

The order of the gimbals from the innermost out is azimuth, inner-roll, pitch and outer-roll. Each gimbal has two synchro transmitters that measure the attitude of the stable element with respect to the gimbal axis. Two servo drive motors, connected in parallel, produce the torque required to position the gimbal through a gear train and to nullify any small amount of backlash which may exist in the precision gearing. The azimuth, pitch, and outer roll gimbals utilize sliprings and brushes for electrical connections; the inner roll gimbal, which is limited to an angular rotation of ±20°, uses flexible leads.

The case of the inertial platform is entirely surrounded by a shroud which permits cooling air to pass around the case. Film-strip heaters, mounted inside the case, are used to bring the platform quickly up to its operating temperature, and to help maintain that temperature.

The small size and low weight of the inertial platform result primarily from the use of Litton's new miniaturized, high-performance inertial components, the gyros and accelerometers. Another contributing size-reduction factor is the removal of all electronics, with the exception of the gyro preamplifiers, from the platform. The two two-degree-of-freedom gyro configuration has permitted an extremely efficient packaging of the stable element; the size of the gimbals and of the platform housing, have therefore, also been appreciably reduced.

G-200 Gyro

The Litton G-200 gyro shown in Figure 2 is three inches in diameter, four inches long, and weighs two pounds. Its 24,000 rpm rotor has an angular momentum greater than 1.5×10^6 cgs units (gm-cm^2/sec). With the gyro spin axis horizontal, the random drift rate about both axes of rotation can be held to less than 0.01 degrees per hour without bias readjustment.

In the exploded view of Figure 3, the major elements of the gyro can be seen. The spherical float, which contains the gyro motor-and-wheel assembly, is mounted in a floated gimbal ring by a pair of jewel and pivot bearings. The ring in turn, is positioned relative to the external case by another pair of jewel and pivot bearings. This configuration achieves the two-degree-of-freedom of the gyro. The float, however, is maintained relatively fixed to the case by a tight servo loop so that 1/4° of mechanical freedom within each axis of rotation is more than adequate.

Each gyro axis has pickoffs to sense any relative motion between its case, which is fixed to the platform, and its float, which at any instant is fixed in inertial space. The pickoff outputs are amplified external to the platform and serve to control the gimbal servo devices to maintain the pickoffs at a null and thereby to keep the platform level.

The pickoffs have a pancake configuration arranged in "push-pull" so that the output is an indication of only rotation of the platform with respect to the float and not translation. The gyro also has torquers on each axis to rotate the float and thereby the platform as a function of earth's rotation and aircraft velocity.

The case surrounds the float with a small clearance so that a relatively small volume of flotation fluid is required. This results in good heat transfer characteristics and low convection current torques. Variations in the volume of the fluid as a function of temperature are accommodated by means of a thin flexible bellows which is extremely pressure sensitive and has a very low hysteresis. The flotation fluid has excellent properties of inertness and stability, and remains liquid to −75°Fahrenheit; so therefore no special storage or transportation problems are involved to maintain the fluid in its liquid state.

The gyro case has steel mounting pads that are lapped to provide a precise surface to mount on a correspondingly precise pad on the platform. The gyros are therefore interchangeable without extensive alignment adjustments.

A-200 Accelerometer

The Litton A-200 accelerometer shown in Figure 4 measures 1.00 inch by 1.135 inches by 1.80 inches, occupies about 2 cubic inches, and weighs seven ounces. Its threshold sensitivity is less than 10^{-5} g and its electrical analog output has an accuracy of better than one part in 50,000. These characteristics have been achieved in an instrument of extreme simplicity - it has only one moving part, the pendulous float. Even in the most severe operational environment, the pendulous float never deflects more than a few seconds of arc from its null position.

In the exploded view of Figure 5, the three subassemblies of the accelerometer (the case, the floated element and the cover) are shown from top to bottom. When assembled, the instrument is hermetically sealed with solder. The pendulum, which is floated in the same type fluid as that employed in the G-200 gyro, is positioned with jewel-and-pivot bearings so that its center of buoyancy lies on the pivot axis. The center of mass however, is displaced from the center of buoyancy so that a known couple is produced whenever the instrument is subjected to an acceleration. Differential transformer-type pickoffs sense any rotations of the float. The external restoring amplifier converts the output of the pickoffs into a d-c current which is applied to torquer coils connected in push-pull and operating in the fields of two permanent magnets located in the cover of the accelerometer. The torquer current which is required to maintain the pickoffs nulled is thus directly proportional to the applied acceleration. This current is applied through a precision resistor and the voltage developed across it represents the output of the instrument. The precision resistor is matched to the individual accelerometer so that the units are electrically interchangeable.

The case of the accelerometer has three mounting pads which, during calibration, are precisely adjusted to be perpendicular to the sensitive axis. The units are therefore also mechanically interchangeable without precision adjustment requirements.

Electronics

The electronics associated with this equipment may be divided into two separate groupings: (1) those platform electronics which stabilize the platform and (2) those which provide the computation functions. Figure 6 shows a block diagram of the electronics required for a four-axis platform stabilization system which utilizes two-degree-of-freedom gyros. The z-axis pickoff signal from one gyro, amplified by an azimuth servo amplifier, actuates the azimuth gimbal torquer. The z-axis signal from the other gyro is used to produce a torque-correction signal at the output of an azimuth caging amplifier. Gyro signals from the x and y axes are resolved through the vehicle heading angle by a resolver on the azimuth gimbal, and the appropriate components of these signals are amplified to drive the pitch and inner-roll gimbal torquers. The outer-roll gimbal torquer is controlled by the inner-roll synchro so as to maintain the azimuth axis perpendicular to the pitch axis: the azimuth, inner-roll and pitch axes are thereby

maintained orthogonal to each other, and gimbal lock cannot occur. The ratio of rotation about the outer-roll axis to the resultant rotation about the inner-roll axis is equal to the cosine of the pitch angle; therefore, the gain of the servo amplifier for the outer-roll gimbal is varied as a function of the pitch angle so as to keep the outer-roll servo loop response uniform.

Depending on the system application, the computation requirements of the inertial systems are satisfied by the use of either analog or digital computers, or by combinations thereof. In the digital approach, the computer performs a number of functions in addition to satisfying the requirements of the inertial system. The full list of computer functions includes the double integration of accelerations, the generation of torquing currents to maintain the platform properly oriented, and the calculations of corrections for ellipticity of the earth, coriolis corrections and navigation coordinate systems. In some systems, the computer includes provisions for the addition of radar doppler-velocity data to both dampen the Schuler oscillation and to give an accurate, in-flight gyro-compass alignment capability. Systems have been designed which have provisions for both manual and automatic insertion of position-fix data from such equipments as TACAN, map matchers, and astrotrackers.

Figure 7 shows the assemblies comprising a current production system, the AN/ASN-28 Navigational Computer Set, which is operational from aircraft carriers in the WF-2 Airborne-Early-Warning Aircraft. This system is a highly accurate, central attitude reference and therefore does not provide a full navigational capability. The units shown are (left to right, top row) the Platform Amplifier, the Platform (including a mount) and the Platform Computer and (left to right, bottom row) the Power Supply, the Erection Controller, the Heading Controller and the Platform Adapter. The main function of the latter assembly is to provide a number of repeated outputs of platform pitch, roll, and heading without loading the platform transducers and degrading their accuracy. The Erection Controller gives the pilot the control of the preflight sequence of the system and the provision to insert necessary initial conditions -- Latitude, Ground Track, and Ground Speed of the aircraft carrier. This same Erection Controller will also permit an in-flight alignment, without doppler data, but with reduced heading-accuracy for emergency requirements. Certain newer systems have provision for the automatic insertion of external velocity data. The Heading Controller provides for the control of the different heading modes of operation.

Figure 8 shows the assemblies comprising the LN-3 Inertial System which is being produced for the F-104G and CF-104 Strike Fighters as their primary navigation system and central attitude reference. Although repackaged to satisfy the specific installation requirements of these aircraft and to incorporate the latest improvements, this system is still a member of the family of systems having the common bond of the P-200 platform. The assemblies shown in the figure are (left to right) the Adapter, the Computer, and the Inertial Platform. The Inertial-Navigation-Control and the Alignment-Control panel are shown on top of the computer. The computer assembly contains the circuitry to perform all the functions of the Platform Amplifier, the Platform Computer, and the Power Supply of the previously described system. The major output of this inertial navigation system is present-position which is supplied to an independent display system.

The LN-3 is an example of the present state of the art of production lightweight inertial systems. Its weight of approximately 80 pounds is offered as evidence to support the statement made in the introduction to this paper, that the smaller weight of current production inertial systems now extends the application of inertial techniques to vehicles and applications in which the weight of inertial systems was formerly a deterrent.

Support Equipment

The systems built around the P-200 platform all have self-test features which provide an automatic go-no-go indication of system readiness. Certain critical signals, such as (1) accelerometer and gyro-pickoff signals and (2) reference and primary voltages, are monitored. Simple diode detectors, clamping circuits, and "and/or" gates are used to check continually these signals and to trigger no-go indicators if any signal disappears or becomes excessive. This approach is quite sensitive to equipment malfunctions, since any component failure within the various closed servo loops will cause the monitored signals to deviate markedly from their nominal values.

Several levels of ground support equipment are available for use with the production inertial systems. The simplest is the Line Test Analyzer shown in Figure 9. This unit is designed for routine upkeep, the daily preparation of aircraft for flight, including preflight and postflight inspections, and trouble shooting. The basic function of the analyzer is the isolation of any unit causing a gross malfunction when the inertial equipment is still installed in the aircraft.

The next level of ground support equipment is typified by the Bench Test Analyzer shown in Figure 10. Briefly this equipment enables an operator to locate a major subassembly or chassis-mounted component which is defective and to test each assembly.

In addition to the special ground support equipments, certain Litton systems have been designed with the capability of being tested by standard, universal military check-out equipments.

The fact that the P-200 platform has been produced in quantity does not imply, of course, that the ultimate design has been achieved. At Litton, one of the major efforts is focused on a military development contract to achieve a still smaller inertial platform, the P-300. This platform, whose mockup is shown in Figure 11, will measure 6-1/2 inches by 6-1/2 inches by 10 inches and will weigh only 13 pounds. The platform has the same basic configuration as the larger, 30-pound, P-200 platform and will equal or better the performance of the P-200.

Until recently the use of inertial systems for airborne application was limited because of the weight involved. However, current availability of production lightweight inertial systems, such as those built around the P-200 platform, has greatly increased the applicability of inertial equipments. Their smaller size and lighter weight permit a wider choice of vehicle types and configurations. The full potential of lightweight inertial systems will be even further increased, however, by the availability of still lighter platforms such as the Litton P-300.

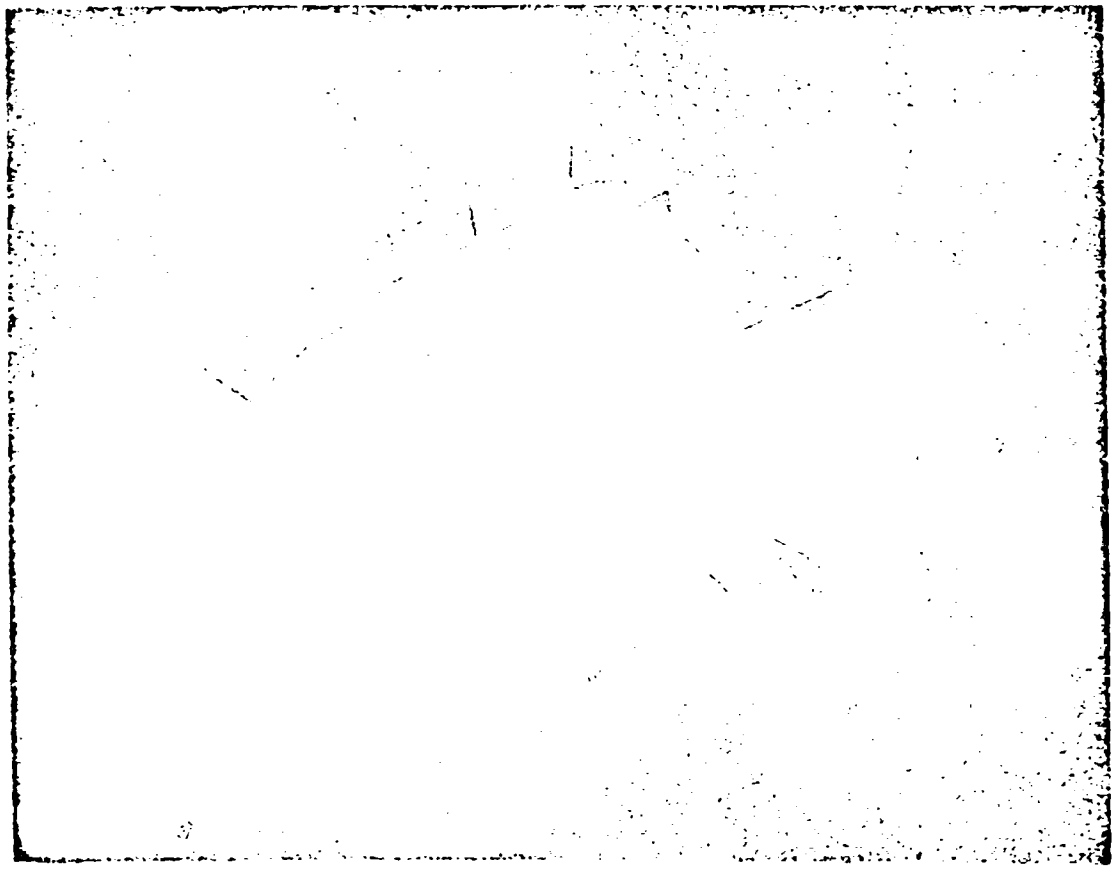


FIGURE 9. P-200 Inertial Platform

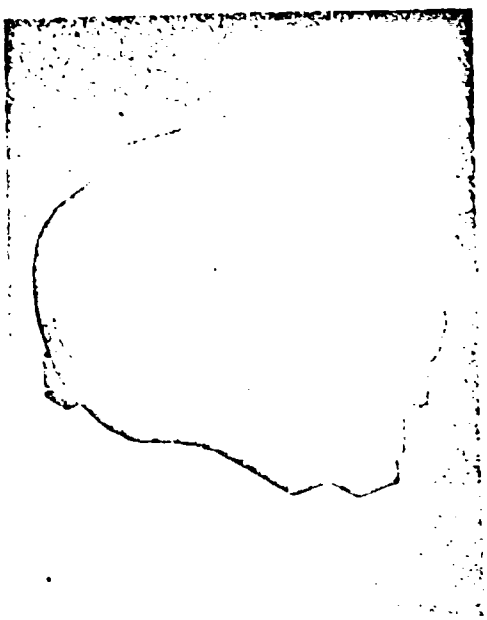


Figure No. 2 Q-200 Quartz

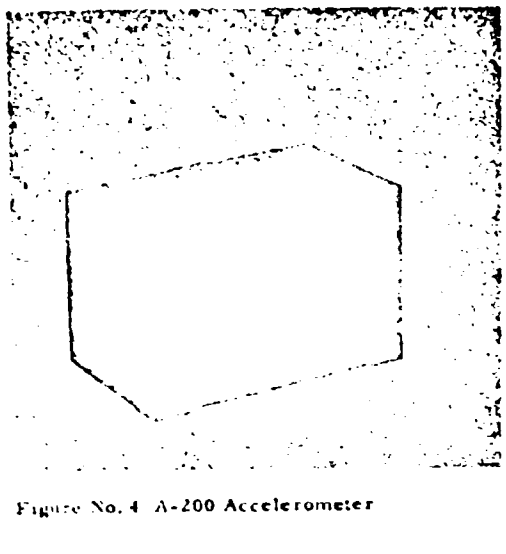
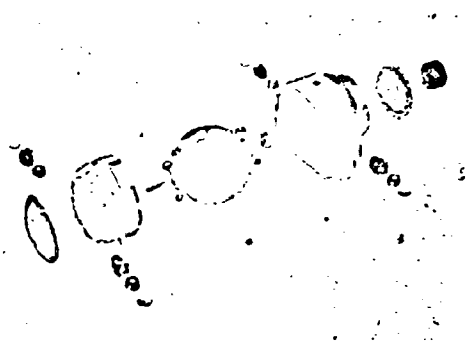


Figure No. 4 A-200 Accelerometer



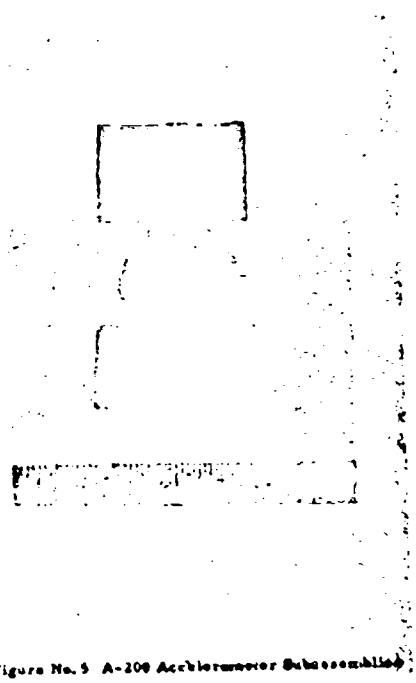


Figure No. 5 A-200 Accelerometer Subassembly

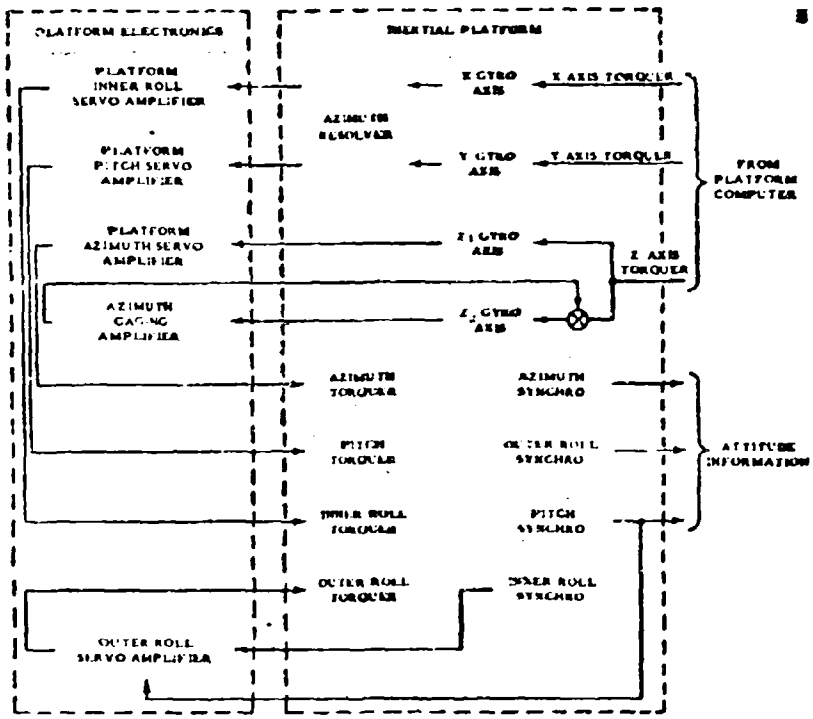
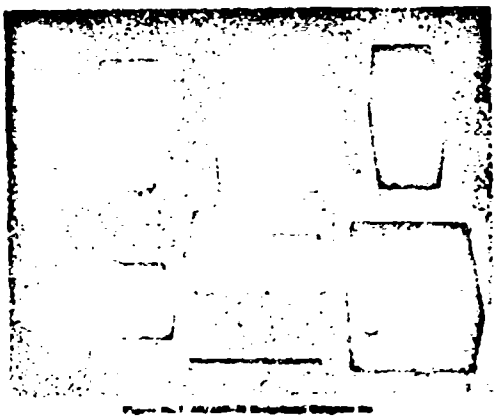
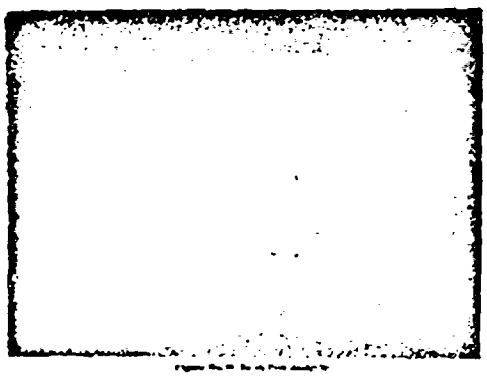
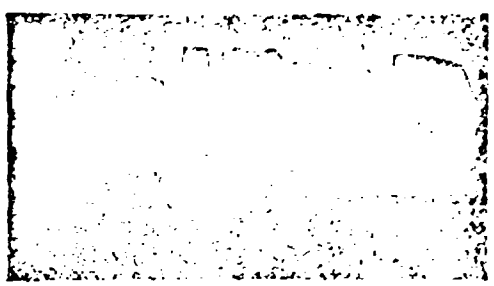
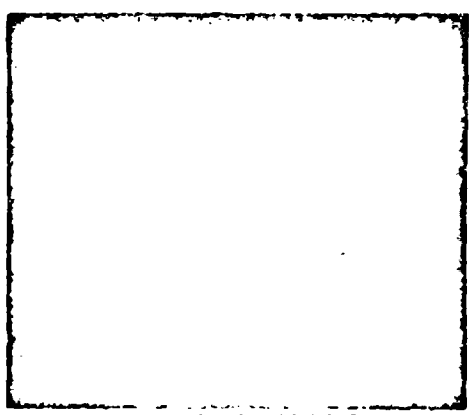


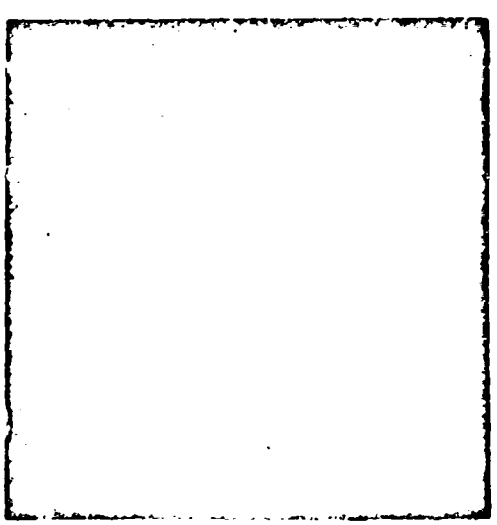
Figure No. 6 Four-Axis Platform Stabilization System











INERTIAL ACCELEROMETERS - THEIR NATURE, CHARACTER, AND LIMITATIONS

By: Martin Maurer, Project Engineer, Kearfott Division, General Precision, Inc.

INTRODUCTION

Frequently the engineer is called upon to synthesize subsystems along the lines of optimization. However, the optimized subsystem or component need not be the best system design.

The engineer's resources are practically unlimited when it comes to utilizing the analytical tools at his command. He usually relies upon the more common tools such as the root locus, Bode, Nyquist and Nichols plots and diagrams as well as the analog computer. In addition, a noise analysis is normally conducted to investigate the basic limitations imposed upon the component or subsystem.

A typical noise analysis would take the following form:

I. List sources of noise pertinent to component

   A. Thermal

      1. electronic derivative circuits
      2. mechanical damping

   B. Fluctuations in pick-off device
   C. Feedback amplifier noise
   D. Variation in magnetic element

II. Determination of magnitude and frequency spectra

      1. Laplace transform of auto-correlation function of noise

III. Determination of optimum filter to maximize the signal-to-noise ratio

This paper discusses two existing types of accelerometers on a component basis. In addition, one of the accelerometers is applied to a set of ideal integrators to yield velocity and distance functions for a typical acceleration profile.

LIST OF SYMBOLS

SYMBOL	DESCRIPTION	UNITS
A	Acceleration	"g" Units
A_c	Cross Acceleration	"g" Units
B	Flux Density of Magnetic Field	Gauss
B	Damping Constant of Accelerometers	Dyne-cm/rad/sec
e_o	Amplifier Output Voltage	Volts
G (s)	Frequency Variant Network	Ohms
I	Accelerometer Restoring Current	Amps
K_1	Servo Gain Constant	Volts/Volt
K_2	Derivative Control Constant	Volts - Sec/Volt
K_a	Amplifier Gain	Volts/Volt
K_E	Signal Generator Sensitivity	Volts/Radian
K_I	Scale Factor of Restoring Coil	Dynes/Amp
L	Laplacian Operator	Sec^{-1}
T	Lead Network Time Constant	Sec
T (s)	Open Loop Transfer Function	(None)
V_o	Signal Generator Output Voltage	Volts
W	Weight of Suspended Mass	Dynes
ω_{b_o}	Open Loop Zero db Crossover Frequency	Rad/Sec
Z(s)	Amplifier Load Impedance	Ohms
g	Acceleration Due to Gravity	"g" Units
k	Electronic Restoring Spring Constant	Dynes
k_1	Spring Constant of Hooke's Joint #1	Dyne-cm/rad
k_2	Spring Constant of Hooke's Joint #2	Dyne-cm/rad
k_s	Spring Constant	Dynes/cm
l	Lever Arm	cm
l_o	Length From Axis of Rotation to Center of Spring	cm
l_w	Total Length of Winding perpendicular to B	cm
m	Mass	Grams
α_L	Lead Network Attenuation Constant	(None)
γ	Accelerometer Pick-Off Angle	Radians
ϵ_q	Threshold Acceleration	"g" Units
v(t)	Velocity	"g" - sec" Units
d(t)	Distance	"g" - sec^2 Units

SPRING-RESTRAINED ACCELEROMETER

The spring-restrained accelerometer is similar in form to a simple pendulum or free pendulum. A free pendulum may be defined as an unrestrained mass of weight W suspended from a fixed point and acted upon by the forces of acceleration and gravity. Physically, the pendulum consists of a small dense mass or "bob" suspended at the end of a very light string.

A restrained pendulum, as distinguished from a free pendulum may be defined as a captured mass of weight W suspended from a fixed point and acted upon by the forces of acceleration and gravity and the forces of restraint.

Definitions applicable in the description of a pendulum, whether they take the form of a free pendulum or a restrained pendulum, are included in the list of symbols.

Since the pendulum, or for that matter an accelerometer, cannot distinguish between gravity or acceleration, it would be proper to define the action of the pendulum under these conditions and combinations of these conditions. The following modes of operation will be analyzed:

1. Free Pendulum

   (a) platform tilted by angle θ,

   (b) platform horizontal with acceleration A,

2. Restrained Pendulum

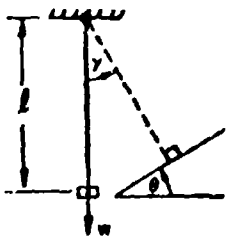
   (a) platform tilted by angle θ

   (b) platform horizontal with acceleration A

   (c) platform tilted by angle θ, with acceleration A,

3. Response of system 2(b) with damping (D) to a step acceleration input.

1. (a) Free pendulum -- platform tilted by angle θ

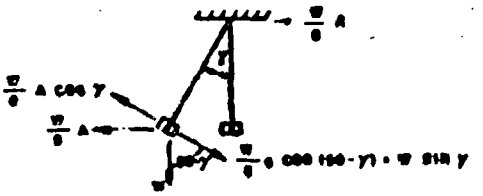


With no acceleration, the free pendulum is always vertical (local) and the tilt of the platform equals the displacement of the pendulum.

$$\gamma = \theta \qquad (1)$$

A transducer is usually employed to detect the angle γ

(b) Free pendulum -- platform horizontal with acceleration A



If we equate the restoring torque equal to the disturbing torque acting on the pendulum, then it can be shown that

$$\tan\gamma = \frac{A}{g} \qquad (2)$$

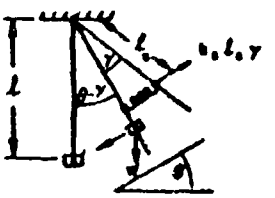
If, γ is small
then, $\tan\gamma \approx \gamma$ (radians)

therefore,

$$\gamma = \frac{A}{g} \text{(radians), for small tilts} \qquad (3)$$

This states that an accelerometer with a sensitivity of 10^{-3}g's would hang-off from local vertical by 10^{-3} radians or approximately 3 arc-minutes.

2. (a) Restrained pendulum -- platform tilted by angle θ



If we equate torques as before,

then,

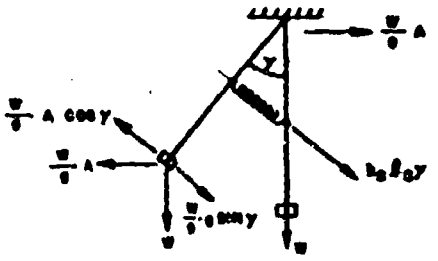
$$k_s I_s^2 \gamma = \frac{W}{g} g \sin(\theta - \gamma) l \qquad (4)$$

if we let $I_s = l$, γ small, and $k_s l \gg W$

then

$$\gamma \approx \frac{W \sin\theta}{k_s l} \text{ (Radians)} \qquad (5)$$

(b) Restrained pendulum -- platform horisontal with acceleration A



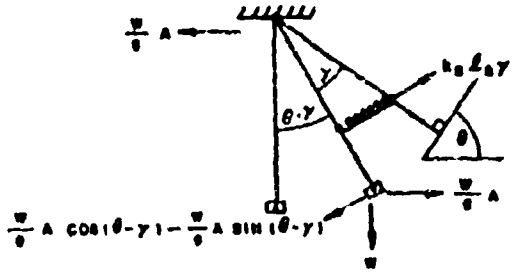
disturbing torques = restoring torque

$$\frac{W}{g}A\cos\gamma\cdot l \approx \frac{W}{g}g\sin\gamma\cdot l + k_s l_s^2\gamma \quad (6)$$

as before, for small γ, and letting $l_s = l$ and $k_s l \gg W$

$$\gamma \approx \frac{WA}{k_s g l} \text{ (radians)} \quad (7)$$

(c) Restrained pendulum -- platform tilted by angle θ with acceleration A



disturbing torque = restoring torque

$$-\frac{W}{g}A\cos(\theta-\gamma)\cdot l + \frac{W}{g}\cdot g\sin$$

$$(\theta-\gamma)\cdot l = k_s l_s^2\gamma \quad (8)$$

For small γ

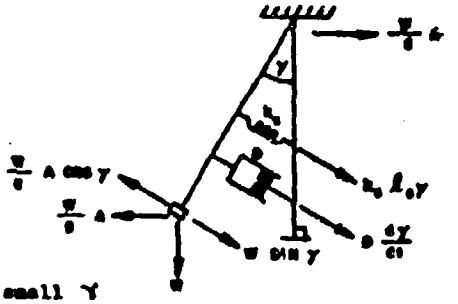
and letting $l_s = l$

$$\gamma \approx \frac{W}{g k_s l}(g\sin\theta - A\cos\theta) \quad (9)$$

It may be shown that a more exact solution is as follows:

$$\gamma = \frac{g\sin\theta - A\cos\theta}{\frac{g\,k_s l}{W} + A\sin\theta + g\cos\theta} \quad (10)$$

3. Response of a damped pendulum to a step acceleration input.



for small γ

$$\frac{W}{g}Al - Wl\gamma \approx \frac{W}{g}\frac{d^2\gamma}{dt^2}l^2 + D\frac{d\gamma}{dt} + k_s l^2\gamma \quad (11)$$

in the Laplace domain A = A(s)

$$\frac{WA(s)}{g} \approx \left[\frac{Wl}{g}s^2 + \frac{D}{l}s + k_s l + W\right]\gamma(s) \quad (12)$$

$$\frac{\gamma(s)}{A(s)} \approx \frac{\frac{W}{g(k_s l + W)}}{\left[\frac{s^2}{\omega_n^2} + \frac{2\zeta}{\omega_n}s + 1\right]} \quad (13)$$

where the undamped natural frequency, ω_n,

$$\omega_n = \sqrt{\frac{g(k_s l + W)}{Wl}} \quad \frac{\text{radians}}{\text{sec}} \quad (14)$$

and the damping ratio ζ,

$$\zeta = \frac{D}{2l}\sqrt{\frac{g}{Wl(k_s l + W)}} \quad (15)$$

if, as in previous cases $k_s l \gg W$

then

$$\omega_n \approx \frac{1}{W}\sqrt{\frac{g k_s}{l}} \quad (16)$$

and

$$\zeta \approx \frac{D}{2l^2}\sqrt{\frac{g}{k_s W}} \quad (17)$$

The inverse transform for a step input of acceleration is as follows:

For,

$k_s \neq 0$

$D \neq 0$

$$\frac{\theta(t)}{|A|} \simeq \frac{W}{g(k_s l + W)} - \frac{\zeta W l}{g}\sqrt{\frac{I}{\gamma I^2 W(k_s l + W) - D^2 g^2}}\, e^{-\frac{D g t}{2\zeta l^2 W}} \sin\left\{\sqrt{\frac{\gamma g l^2 W(k_s l + W) - D^2 g^2}{2 I^2 W}}\, t + \cos^{-1}\frac{D}{2\zeta l}\sqrt{\frac{g}{W l(k_s l + W)}}\right\} \tag{18}$$

For,

$k_s = 0$

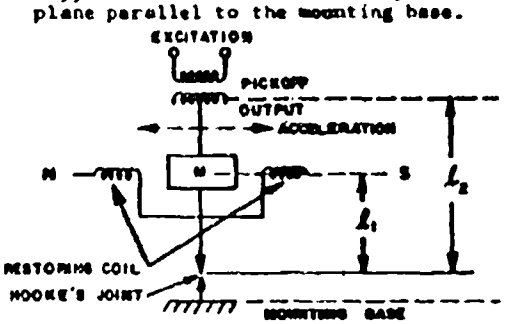
$D \neq 0$

$$\frac{\theta(t)}{A} \simeq \frac{1}{g} - \frac{\zeta W l}{g}\sqrt{\frac{I}{\gamma I^2 W^2 - D^2 g}}\, e^{-\frac{D g t}{2\zeta l^2 W}} \sin\left\{\sqrt{\frac{\gamma g W^2 l^2 - D^2 g^2}{2 I^2 W}}\, t + \cos^{-1}\frac{D}{2\zeta W l}\sqrt{\frac{g}{l}}\right\} \tag{19}$$

PENDULUM-TYPE ACCELEROMETER

Definition and Properties

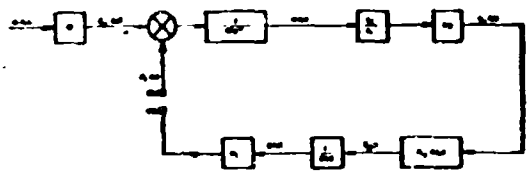
The pendulous accelerometer in this report is of the inverted pendulum type using a Hooke's joint for pendulum support so that it moves freely in a plane parallel to the mounting base.



The above figure represents one axis of a pendulous accelerometer.

The electrical pick-off on its sensing axis detects displacement of the pendulum from its unaccelerated position. The output of the E-bridge pick-off is amplified, demodulated, and fed back to the restoring coil. The magnitude and polarity of the current in the restoring coil is a measure of the acceleration along that axis.

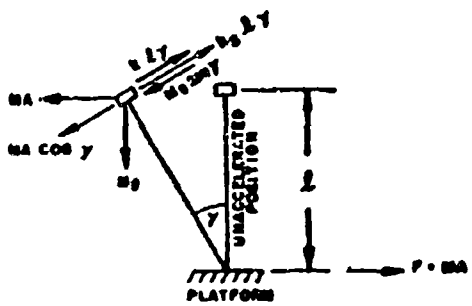
A simplified block diagram of the accelerometer loop is seen below. The parameters appear in the list of symbols.



Gravity and Spring Force

Since the accelerometer loop is very tightly captured, one may consider the effects of gravity and spring constant as negligible.

This is illustrated in the simple force-mass diagram.



The force-mass diagram shows a captured pendulum mounted on a horizontal platform and subjected to an acceleration A.

where,

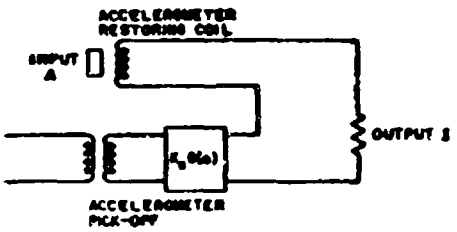
$k l \gamma$ = restoring force due to electronic feedback

$k_s l \gamma$ = spring force

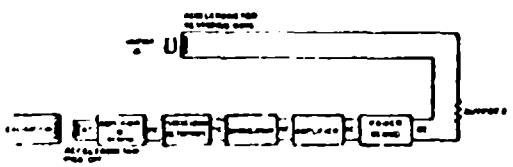
Since the disturbing torque equals the restoring torque for the force mass system it may be shown that for small values of γ .

$$\gamma = \frac{MA}{kl + (k_s l - Mg)} \tag{20}$$

In the previous equation Mg is proportional to the force due to gravity, $k\ell$ is proportional to the electronic restraining force resulting from the accelerometer amplifier feedback loop, and $k_s\ell$ is proportional to the spring force.



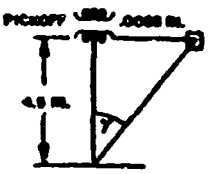
ACCELEROMETER LOOP



A TYPICAL ACCELEROMETER CLOSED LOOP

The spring force and the force of gravity have opposing effects on the accelerometer. An adjustment procedure nearly has these opposing forces cancelling one another with the result that their combined effect approaches zero.

For a constant acceleration input of 10g, the pick-off output voltage is 60 MV. For a constant 1g input the pick-off voltage is approximately 6 MV. Since the pick-off has a sensitivity of 20 MV/.001 inches, this represents .0003 inches hang-off of the pendulum from its unaccelerated position.



From the above diagram, it can be seen that $\gamma = \frac{.0003}{4.5} = \frac{1}{15,000}$ radians for a constant input of 1g acceleration. This is equivalent to 14 arc-seconds hang-off. At this extremely low hang-off angle the effects of gravity and spring force have negligible effect.

Open Loop Transfer Function

The open loop transfer function (T) may be obtained from the accelerometer block diagram.

$$T(s) = \frac{\frac{K_2 K_a K_i}{2} \frac{\ell_a}{L_i} G(s)}{MS^2} \qquad (21)$$

Using typical values

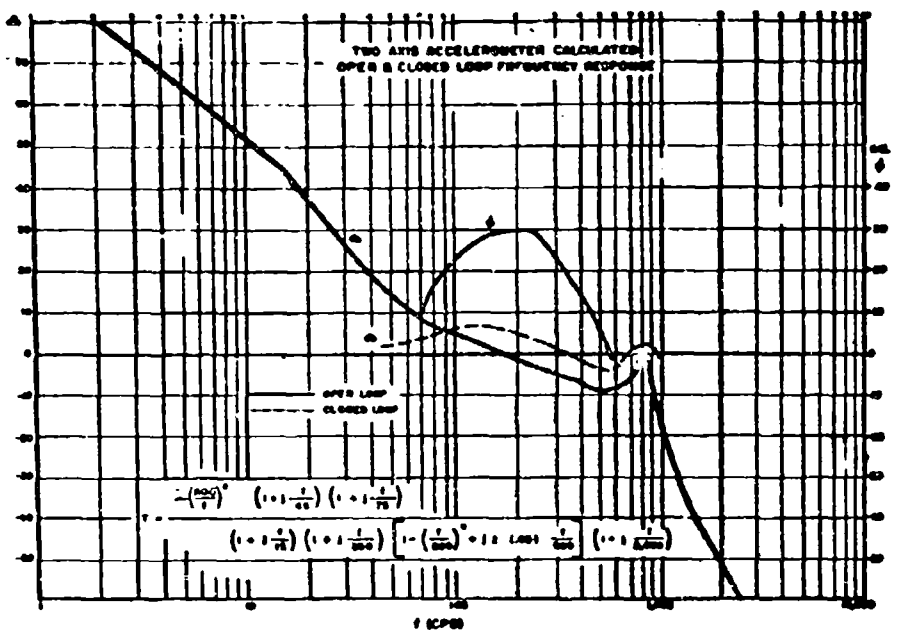
$$T(j\omega) = -\left[\frac{2\pi \times 206}{\omega}\right]^2 \qquad (22)$$

The natural frequency of the above loop

$$f_o = 206 \text{ cps.}$$

A decibel versus frequency plot of T (jf) will pass through 0 db at 206 cps with a -12 db/octave slope indicating instability.

Since the pendulum exhibits an 800 cps mechanical resonance, it is advisable to have a low open-loop gain at this frequency. Therefore, a lag-lead network may be used to stabilize the loop and slightly decrease the bandwidth of the system. The use of the lag network by itself would tend to isolate the closed loop frequency response peak magnitude from the 800 cycle resonant peak. By itself the lag network is not adequate for stabilization, therefore the need for a lag-lead network. Also the lag network does provide a slight reduction in the accelerometer hang-off. A



large reduction of gain at low frequencies caused by the lag network would be detrimental to the system performance since the accelerometer may not be able to capture the pendulum for high g inputs at the low frequencies. For this reason a 3:1 reduction representing 9.5 db was chosen for the lag network.

The function $G(\omega)$ is of the following type:

$$G(\omega) = \frac{\left(1 + \frac{s}{282}\right)\left(1 + \frac{s}{470}\right)}{\left(1 + \frac{s}{942}\right)\left(1 + \frac{s}{2,580}\right)} \qquad (23)$$

The open loop transfer function with lag-lead stabilization is as follows:

$$T(jf) = \frac{\left(\frac{206}{f}\right)^3\left(1 + j\frac{f}{95}\right)\left(1 + j\frac{f}{75}\right)}{\left(1 + j\frac{f}{18}\right)\left(1 + j\frac{f}{380}\right)\left[1 - \left(\frac{f}{800}\right)^2 + j2(.05)\frac{f}{800}\right]} \qquad (24)$$

The above Bode plot represents the calculated open-and-closed loop frequency response of the accelerometer. The curves indicate that the effect of the lag network reduced the 0 db cross-over frequency. The compensated system illustrated is stable with a phase margin of + 30 degrees.

ACCELEROMETER LIMITATIONS

This paper will discuss three basic limitations of the accelerometer. The following limitations are discussed in a current research project:

1. Orientation of input axis to mechanical elements and sensitive axis (including orthogonality of two-axis units.)
*2. Threshold sensitivity
*3. Scale factor accuracy and stability
4. Bias error
5. Damping factor
6. Linearity
7. Cross coupling
8. Nonlinearity of electronics
9. Restraints
10. Vibration rectification
*11. Transient response

*Discussed in this paper (items 2 and 3 were prepared with the assistance of S. Dardarian, Project Engineer, Kearfott Division of General Precision, Inc.)

12. Saturation

13. Hysteresis (zero uncertainty)

14. Cross axis vibration effects

15. Dynamic range

THRESHOLD SENSITIVITY

a. Spring Restrained:

The pivots have a static frictional torque

T_f

$T_f = Fl$ at breakaway

$F = ma_T 980$, a_T = threshold acceleration

$$\therefore a_T = \frac{T_f}{ml(980)} \qquad (25)$$

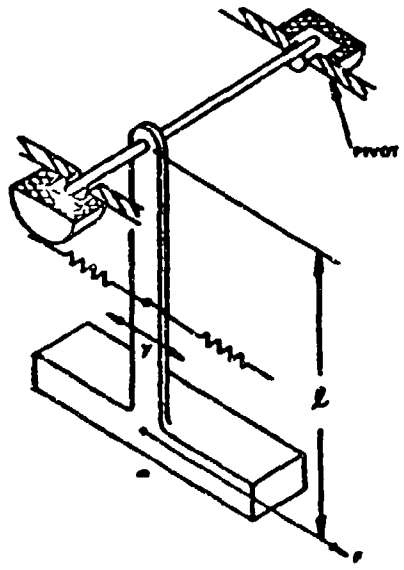
for well machined bearings

$T_f \approx 20$ dyne cm

$ml \approx 6 \times 3 = 18$ gm-cm

$$\therefore a_T \approx \frac{20 \text{ dyne cm}}{18\ (980)}$$

$a_T \approx 1 \times 10^{-3}$ g's threshold acceleration.



Forces arising from displacement of spring are considered negligible since the definition of threshold stipulates small motions of the pendulum.

b. Pendulous Accelerometer

The threshold sensitivity of the inverted pendulous accelerometer depends upon how much hysteresis the materials of construction have, how sharply the take-off null is defined, the threshold sensitivity of the amplifier, and the restraints acting on the armature. Ideally a net spring torque does not exist with a 1g cross acceleration. The Hooke's joint suspension can not be considered a restraint. Small secondary effects arise due to small mismatches of the spring rate to the mass of the pendulum.

SCALE FACTOR ACCURACY AND STABILITY

a. Spring-Restrained Accelerometer.

$$\frac{V_o}{A} = \frac{980 K_g n l \cos\alpha}{I^2 k_s + mgl} \qquad (26)$$

The scale factor V_o/A is affected by:

1) error in l due to non-homogeneity of the material

2) error in l due to machining.

3) error in m due to variation in density of material

4) error in m due to non-homogeneity of material

5) error in l_s due to machining

6) variation in k_s due to changing of spring dimension (mechanical errors)

7) variation K_g due to changes in air gap

8) variation in K_g due to winding errors

9) variation in K_g due to variation in magnetic qualities of the materials.

10) local g changes.

The scale factor is set for the instruments individually by selectively choosing over or undersized springs to compensate for composite errors. Generally, the scale factor may be set to within ± 2%.

Scale factor variations occur when,

1) ℓ varies due to temperature changes

2) $m\ell$ varies due to mass shifts i.e. center line of rotation changes

3) ℓ_s varies due to temperature

4) k_s changes due to temperature and aging

5) K_g varies due to temperature change

b. Pendulous Accelerometer

assuming all errors equal zero, we have

$$A_1 = \frac{I_1}{980m}\left[\frac{k_1 i_1}{K_g K_a} + K_i i_1 - \frac{I_1 i_1}{K_g K_a} m A_2\, 980\right] \qquad (27)$$

and if $k_1 = 980\, m\ell_1$

$$\frac{A_1}{i_1} = \frac{1}{I_1 m\, 980}\left[\frac{k_1}{K_g K_a} + K_i - \frac{k_1 A_2}{K_g K_a}\right] \qquad (28)$$

$$\frac{A_1}{i_1} = \frac{K_i}{980 m I_1} + \frac{k_1}{980\, K_g K_a m I_1}\left[1 - A_2\right] \qquad (29)$$

similarly

$$\frac{A_2}{i_2} = \frac{K_i}{980 m I_2} + \frac{k_2}{980\, K_g K_a m I_2}\left[1 - A_1\right] \qquad (30)$$

Since the expressions are similar, only one axis need be discussed. In general,

$$\frac{A}{i} = \frac{K_i}{980 m \ell} + \frac{k}{980\, m \ell K_g K_a}\left[1 - A_2\right] \qquad (31)$$

The scale factor is affected by K_i to a very large extent, K_g and K_a by a smaller amount if the cross acceleration is approximately 1g.

$$F\ell = \frac{B \ell_w i \ell}{10} \qquad (32)$$

$$\frac{F\ell}{i} = K_i = \frac{B \ell_w \ell}{10} \qquad (33)$$

In general,

$$\frac{A}{i} = \frac{B\ell_w}{m\, 9800} + \frac{k}{980\, K_g K_a m \ell}\left[1 - A_2\right] \qquad (34)$$

Scale factor variations (fixed)

B may vary due to:

1) porosity of magnetic material
2) size of the air gap
3) permeability of the material
4) size of temperature-compensating shunt

ℓ_w may vary due to:

1) size of spool (since the number of turns is counted)

2) change in packing factor of winding causing different average diameter

g must be known accurately

m - mass may vary due to:

1) machining tolerances
2) density variations
3) amount of cement used
4) amount of wire on coils
5) amount of solder on terminals

k may vary due to:

1) variation of modulus of elasticity of spring material

2) variation in dimensions of Hooke's joint

ℓ may vary due to:

1) machining tolerances (accumulative)
2) variation in densities of materials

K_g may vary due to:

1) variation in number of turns in primary

2) variation in number of turns in secondary

3) dimensional tolerances

4) angular positioning of stationary coils

5) angular positioning of moving coils

6) variation of input voltage

7) variation of input frequency

8) variation in the input impedance of the amplifier

Variation in scale factor (time-dependent)

a) B may vary due to

1) temperature effects due to improper compensation

2) aging of magnet
3) temperature effects on scale factor adjusting screw

b) ℓ_g may vary due to

1) temperature effects changing length approximately

$$1 \times 10^{-3}\ \%/\ ^\circ F$$

2) Joule heating effect (I^2R losses) affects the unit 1×10^{-3} %/°F

c) mass may vary due to

1) mass shifts due to non-uniform expansion of parts

2) oxidation of materials

d) k may vary due to

1) variation of modulus of elasticity due to temperature

2) variation of dimensions due to temperature

3) work hardening of material under repeated stressing

e) ℓ may vary due to

1) growth of material due to instabilities

2) change in length due to temperature

f) K_g may vary due to

1) variation in dimensions changing flux coupling coefficient

2) variation in resistance of primary windings

It is difficult to assign numbers to these errors for they are dependent upon how the accelerometer machined parts are detailed. The factors which are not time dependent are compensated for by a small shunt screw which varies the flux density. The other factors which are time-dependent are predominantly due to heat-treating the machining parts or dependent upon temperature effects. These variations are so complex that any attempt at analysis would be too involved to have any meaning. The best approach here is to use data supplied by the manufacturer.

The effect of non-uniform heating is to tend toward large errors. This must be avoided in order to obtain optimum performance. Also, since the temperature-compensation occurs over a given operating range, the unit's environment should be thermostated to maintain this temperature.

TRANSIENT RESPONSE

We noticed earlier that the uncompensated open-loop transfer function of the inverted-pendulum type of accelerometer has the following form:

$$T(j\omega) = -\left(\frac{2\zeta\omega_n + 2\alpha}{\omega}\right)^2 \qquad (21)$$

In place of the lag-lead function associated with G(s), a proportional plus derivative network was chosen mainly for simplicity. It follows that:

$$T(j\omega) = \frac{K_g K_m K_1 \frac{\ell_s}{\ell_i} K_2 \left(\frac{K_1}{K_2} + j\omega\right)}{Ms^2} \qquad (35)$$

This results in derivative control for stability and does not overload the analysis with a higher order characteristic equation. The resultant closed-loop transfer function relating output current to input acceleration is as follows:

$$\frac{I(s)}{a(s)} = \frac{C(s+\alpha)}{s^2 + 2\zeta\omega_n s + \omega_n^2} \qquad (36)$$

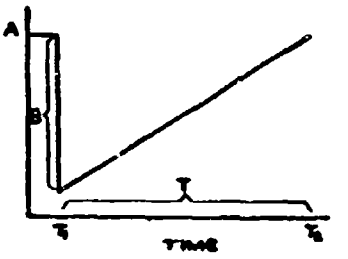
Where the denominator is of the classic second order form, C equals the system scale factor and $\frac{K_1}{K_2}$ equals the ratio of α. Equation 36 is also equivalent to the form,

$$\frac{I(s)}{C\,a(s)} = \frac{(s+\alpha)}{(s+\alpha)(s+\gamma)} \qquad (37)$$

α and γ being equivalent to the pair of complex roots $\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$

Now, if we connect our accelerometer to a set of ideal integrators and subject it to an input, it would be possible to obtain the transient response on an acceleration, velocity and distance level.

The following acceleration profile was assumed:



$$a(t)=Au(t)-Bu(t-\tau_1)+\frac{B}{T}(t-\tau_1)u(t-\tau_1) \qquad (38)$$

In the Laplace domain this equation takes the form,

$$a(s)=\frac{A}{s}-\frac{B}{s}\varepsilon^{-s\tau_1}+\frac{B}{Ts^2}\varepsilon^{-s\tau_1} \qquad (39)$$

Therefore:

$$\frac{I(s)}{C}=\frac{(s+a_0)}{(s+\alpha)(s+\gamma)}\left[\frac{A}{s}-\frac{B}{s}\varepsilon^{-s\tau_1}+\frac{B\varepsilon^{-s\tau_1}}{Ts^2}\right] \qquad (40)$$

The accelerometer response follows:

$$\frac{I(t)}{C}=A\left[\frac{a_0}{\alpha\gamma}+\frac{(a_0-\alpha)}{\alpha(\alpha-\gamma)}\varepsilon^{-\alpha t}+\frac{(a_0-\gamma)}{\gamma(\gamma-\alpha)}\varepsilon^{-\gamma t}\right]$$
$$-Bu(t-\tau_1)\left[\frac{a_0}{\alpha\gamma}+\frac{(a_0-\alpha)}{\alpha(\alpha-\gamma)}\varepsilon^{-\alpha(t-\tau_1)}+\frac{(a_0-\gamma)}{\gamma(\gamma-\alpha)}\varepsilon^{-\gamma(t-\tau_1)}\right] \qquad (41)$$
$$+\frac{B}{T}u(t-\tau_1)\left[\frac{(a_0-\alpha)}{\alpha^2(\gamma-\alpha)}\varepsilon^{-\alpha(t-\tau_1)}+\frac{(a_0-\gamma)}{\gamma^2(\alpha-\gamma)}\varepsilon^{-\gamma(t-\tau_1)}\right.$$
$$\left.+\frac{a_0}{\alpha\gamma}(t-\tau_1)\right]$$

To obtain velocity, the expression is multiplied by $\frac{1}{s}$

$$\frac{I(s)}{sC}=\frac{(s+a_0)}{(s+\alpha)(s+\gamma)}\left[\frac{A}{s^2}-\frac{B\varepsilon^{-s\tau_1}}{s^2}+\frac{B\varepsilon^{-s\tau_1}}{Ts^2}\right] \qquad (42)$$

$$\frac{v(t)}{C}=A\left[\frac{(a_0-\alpha)}{\alpha^2(\gamma-\alpha)}\varepsilon^{-\alpha t}-\frac{(a_0-\alpha)}{\gamma^2(\gamma-\alpha)}\varepsilon^{-\gamma t}+\frac{a_0t}{\alpha\gamma}+\frac{\alpha\gamma-a_0(\alpha+\gamma)}{\alpha^2\gamma^2}\right]$$
$$-Bu(t-\tau_1)\left[\frac{(a_0-\alpha)}{\alpha^2(\gamma-\alpha)}\varepsilon^{-\alpha(t-\tau_1)}-\frac{(a_0-\gamma)}{\gamma^2(\gamma-\alpha)}\varepsilon^{-\gamma(t-\tau_1)}+\frac{a_0(t-\tau_1)}{\alpha\gamma}\right.$$
$$\left.+\frac{\alpha\gamma-a_0(\alpha+\gamma)}{\alpha^2\gamma^2}\right] \qquad (43)$$
$$+\frac{Bu}{T}(t-\tau_1)\left[-\frac{(a_0-\alpha)}{\alpha^2(\gamma-\alpha)}\varepsilon^{-\alpha(t-\tau_1)}+\frac{(a_0-\gamma)}{\gamma^2(\gamma-\alpha)}\varepsilon^{-\gamma(t-\tau_1)}\right.$$
$$\left.+\frac{a_0(t-\tau_1)^2}{2\alpha\gamma}+\frac{\alpha\gamma-a_0(\alpha+\gamma)}{\alpha^2\gamma^2}(t-\tau_1)-\frac{a_0}{\alpha^2\gamma^2}\right]$$

To obtain distance, the original expression is multiplied by $\frac{1}{s^2}$.

$$\frac{I}{s^2C}=\frac{(s+a_0)}{(s+\alpha)(s+\gamma)}\left[\frac{A}{s^3}-\frac{B\varepsilon^{-s\tau_1}}{s^3}+\frac{B\varepsilon^{-s\tau_1}}{Ts^4}\right] \qquad (44)$$

$$\frac{d(t)}{C}=A\left[\frac{(a_0-\alpha)\varepsilon^{-\alpha t}}{\alpha^2(\gamma-\alpha)}+\frac{(a_0-\gamma)\varepsilon^{-\gamma t}}{\gamma^2(\gamma-\alpha)}+\frac{a_0}{2\alpha\gamma}t^2-\frac{a_0}{\alpha^2\gamma^2}\right]$$
$$-Bu(t-\tau_1)\left[-\frac{(a_0-\alpha)\varepsilon^{-\alpha(t-\tau_1)}}{\alpha^2(\gamma-\alpha)}+\frac{(a_0-\gamma)\varepsilon^{-\gamma(t-\tau_1)}}{\gamma^2(\gamma-\alpha)}\right.$$
$$\left.+\frac{a_0(t-\tau_1)^2}{2\alpha\gamma}+\frac{\alpha\gamma-a_0(\alpha+\gamma)}{\alpha^2\gamma^2}(t-\tau_1)-\frac{a_0}{\alpha^2\gamma^2}\right] \qquad (45)$$
$$+\frac{Bu(t-\tau_1)}{T}\left[\frac{(a_0-\alpha)\varepsilon^{-\alpha(t-\tau_1)}}{\alpha^2(\gamma-\alpha)}-\frac{(a_0-\gamma)\varepsilon^{-\gamma(t-\tau_1)}}{\gamma^2(\gamma-\alpha)}+\frac{a_0(t-\tau_1)^3}{6\alpha\gamma}\right.$$
$$\left.+\frac{\alpha\gamma-a_0(\alpha+\gamma)}{2\alpha^2\gamma^2}(t-\tau_1)^2-\frac{a_0t}{\alpha^2\gamma^2}(t-\tau_1)-\frac{1}{\alpha^2\gamma^2}\right]$$

In each of the above equations, the term $\frac{\alpha\gamma-a_0(\alpha+\gamma)}{\alpha^2\gamma^2}$ can be made zero by selecting a value for a_0 equal to $\frac{\alpha\gamma}{\alpha+\gamma}$.

This value turns out to be equal to $\frac{\omega_n}{2\zeta}$

If we assume that our acceleration profile is of several seconds duration, we can ignore the exponential terms since they are damped sinusoids with time constants of millisecond duration. The last term of the velocity expression $\frac{a_0}{\alpha^2\gamma^2}$ is equivalent to $\frac{1}{2\zeta\omega_n^2}$ and its contribution to velocity or distance errors is exceedingly small.

It can be concluded, therefore, that the transient response for the chosen acceleration profile was too rapid for any errors to enter the system. In addition, if the time constant of the acceleration profile was comparable to, or faster than, the accelerometer response, errors might well exist due to the exponential terms. In that particular situation it is conceivable that a compensation method would be used to cancel out the lag effects of the accelerometer.

NOTES

1. This work was done at Kearfott Division of General Precision, Inc. for the Aeronautical Research Laboratories of A.R.D.C. Contract No. AF 33(616)-6502

2. Liu, F.F., and Berwin, T.W., "Extending Transducer Transient Response by Electronic Compensation for High Speed Physical Measurements" Review of Scientific Instruments Volume 29 Number 1, January 1958, Page 14.

JAMMING EFFECTIVENESS INSTRUMENTATION

By: Captain C. B. Redwine and Mr. C. H. Meyer, ECM Test Branch, Rome Air Development Center

In this paper the authors are privileged to report on a program representing the joint efforts of Rome Air Development Center and Vitro Laboratories of West Orange, N.J.

Introduction

The ECM Engineering Test Facility of the Rome Air Development Center has the primary mission of performing engineering evaluations of USAF developmental ECM systems and subsystems. Our instrumentation is based upon the philosophy that engineering tests should provide quantitative, repeatable, data to engineers responsible for the development of ECM techniques and equipment. This data is essential in evaluating breadboard models, results of modifications, and in making technical comparisons of competitive equipment. As a result of experience gained through performance of this mission it has become apparent that current techniques in ECM Engineering testing suffer from two serious deficiencies:

1. There is no general agreement upon the definition of Jamming Effectiveness.

2. Current measurement techniques are almost entirely based upon human evaluations. As a result, quantitative precision, sensitivity, repeatability, and dynamic range are virtually non-existant.

OBJECTIVE

As a first step toward overcoming these deficiencies we have defined four specific cases of Jamming Effectiveness and have identified a measurable parameter for each case. Breadboard model equipment has also been developed and tested for the automatic measurement of each parameter.

The overall objective in the design of this equipment was to instrument the radars of the ECM Test Facility to provide automatic measurement with outputs in a form compatible with the Automatic Monitoring and Data Processing System of the facility.

The instrumentation developed to measure these parameters is officially designated Countermeasures Evaluation Set AN/GLQ-7(XW-1), but is more easily referred to in this paper as the JE Meter. A separate subset is provided for each type of measurement. The subsets are referred to as Self-Screening, Area Screening, Signal to Noise and Track Comparator.

A Gating Group, AN/GPA-90(XW-1), equipment required for operation of the JE Meter was developed concurrently. Space does not permit a technical description of the Gating Group other than to say it delivers precision gates corresponding to the range and azimuth of targets tracked by the AN/MSQ-1A. These gates are indicated in the Self Screening and Signal to Noise Subsets.

The specific cases defined and parameters to be measured are:

1. Self Screening effectiveness is a measure of the ability of a jamming aircraft to conceal itself. The parameter chosen to measure this effectiveness is the ratio of Jamming to Jamming plus Echo.

2. Area Screening effectiveness is a measure of the ability of a jamming aircraft to conceal other aircraft. The parameter chosen to measure effectiveness is Reduction of Radar Detection Area.

3. Deception Device effectiveness is a measure of the ability of that device to deny tracking information. Two parameters were chosen according to the type of intended victim radar. For auto-track victims we measure Track Error. For search or height finder victims the measurement is Signal to Noise Ratio.

Equipment and Results

In the equipment descriptions which follow, details of the circuitry are omitted because they would be quite familiar to workers in the fields of radar and ECM. Instead of circuitry, emphasis here is placed on the functional schemes utilized and its applicability to ECM on the basis of results obtained.

Since the Self Screening subset incorporates ideas common to two other units we will discuss it here first to avoid repetition and facilitate the understanding of the other subsets.

Self Screening:

To measure the self screening effectiveness of an airborne jammer, the radar video signal is gated over an area which includes the target echo and is then compared to a signal proportional to the jamming amplitude. Echo signals strong enough to exceed the jamming level are sampled and recorded as the number of hits in one scan of the target by the antenna. Since the number of allowable sweeps in the scan is preset, the hit-sweep ratio can be determined.

The photograph (fig. 1) shows the physical configuration.

Figure 2 is a simplified block diagram of the self screening instrumentation. Radar video information is extracted prior to the video clipper stage of the radar set. This is

necessary to provide sensitivity and dynamic range greater than the PPI presentation. A sample of the noise is taken during each radar sweep by means of a preset gate which is established during the radar dead time. This gated video sample is integrated, boxcared and amplified. This boxcared and amplified signal is proportional to the noise amplitude of the jamming signal alone and has constant amplitude for the interpulse period of the radar. This signal establishes a threshold level for comparing the video sample taken in a gate containing the aircraft target return.

To sample the target return, a range gate is established by means of a range trigger generated from a radar tracking the same target. This trigger is time synchronized to the search radar and is adjusted to establish a precise range gate which is comparable in width to the search radar pulse. This gated video information is negatively added to the threshold level established by the jamming sample and compared to a dc clip-level signal. Any excursion of the video sample beyond the clip level is amplified and triggers a single pulse generator which causes a counter to advance by one count.

Azimuth information is also obtained from the tracking radar. This information is synchronized to the search radar antenna, and the width of the azimuth gate is established knowing the PRF, antenna speed, and the beamwidth of the search radar antenna. Counts are established for the number of radar sweeps per beamwidth and also for the number of sweeps per beamwidth wherein the video sample exceeds the threshold. These two counts provide the ratio of hits per scan in percent.

Flight test results showing correlation between machine data and operator Blip/Scan data have not been conclusive due to the use of interim gating equipment to provide azimuth gates around the target. Errors up to ± 2.5 degrees occurred which were equivalent to the 3 db antenna beamwidth of the radar under test. However, simulated target and jamming signals have been used, and the results shown in figs. 3 and 4 demonstrate the ability of the machines to perform their intended function. Figure 3 is the detection calibration curve achieved by injecting a variable amplitude rf pulse into the radar and recording the selfscreening count for each level of the rf signal. It is significant to note that operator MDS level is approximately a 50% count and the machine achieves at least 3 db more sensitivity than an operator. Figure 4 shows the effects of three different noise levels on the detection calibration curve. The three curves were taken at receiver noise level and at 6 and 12 db above receiver noise. The three curves are displaced approximately 6 db apart and have a symmetrical shape, indicating the detection system is relatively insensitive to the amplitude of the signals being detected.

Area Screening

The area screening portion borrows several techniques already described in the self screening unit. The measurement performed is actually a counting of artificial targets in the presence of the jamming on a per antenna revolution basis. Five target sizes are injected cyclically, one per range scan, at the end of each scan. For a typical radar with antenna speed of 5 RPM and 400 pulses per second there would be 4800 targets counted with no jamming. Jamming, of course, reduces the count. The number of "looks per revolution" is called No and the number of detections is N. The ratio of N to No is a measure of the useful radar coverage and is presented on counters. The same data is presented on a panel meter calibrated in percentage screening of the normal radar coverage area. The artificial target sizes are adjustable; however, they are ordinarily set so that the smallest target simulates the echo from the test aircraft at maximum radar range with the four others increasing in size to represent the same aircraft at shorter ranges corresponding to increments of equal coverage area.

Before passing to the block diagram, refer to the timing chart (fig. 5) which provides an easier understanding of the general operational scheme. Only the essential elements are shown. Lines 1 & 2 show the antenna rotation signals and continuous radar triggers. The lower 18 lines are an expanded view of conditions during a period of 6 radar triggers.

The simplified block diagram (fig. 6) shows more clearly the way in which synthetic targets are injected into the radar and the way radar noise plus synthetic targets are returned to the AE meter for sampling and counting.

An indication of actual performance is shown in fig. 7 and 8. The data for the calibration curve was obtained by adjusting the S-level synthetic signal to provide a 95 percent count with normal receiver noise and then injecting external noise into the radar. The curve shows a useful range of about 32 db. The machine count is the average of 5 antenna sweeps. The reason for the minimum count of 20% is that with such large amounts of noise, cumulative instabilities produce counts from the noise itself. Figure 8 showing the correlation between synthetic target count and useful radar area was obtained using normal radar operator procedures based upon the familiar scheme of "Condition Echo" one through five.

Signal to Noise:

Figure 9 is a simplified block diagram of the signal to noise ratio measuring instrumentation. The output of the radar is obtained at the radar preamplifier. This output is fed to an IF amplifier and detector with bandwidth identical to the radar set being evaluated. These circuits

are required to bypass the action of the radar AGC. A separate AGC stage is utilized to gate out the initial transmitter pulse and ground clutter signals to provide an IF signal that contains only radar noise plus the target signal of interest. A dc voltage is developed to bias the IF amplifier gain in order to keep the average noise output at a constant level. A target sample gate is established which is identical to that created in the self screening instrumentation. The output of this range-azimuth video gate consists of a constant level of noise plus the target signal of interest. A video sampler accepts this gated video signal and produces a sawtooth signal which activates a pulse counter. The total number of pulses fed to the pulse counter is proportional to signal strength plus noise to noise. The output consists of visual numeric counters for manual or photographic recording and a special output which is fed to a punched paper tape recorder for use in the computer.

The primary use of this machine was to evaluate the effectiveness of airborne decoy systems. The signal strengths of both radar returns would be compared to evaluate the performance of the decoy. Experimental tests have shown several other uses for this instrument. Figure 10 is a plot of the horizontal pattern of the radar under test. A remote signal generator was synchronized with the radar trigger and the signal amplitude was measured at 1 degree azimuth intervals as the antenna was rotating. Figure 11 shows the effects of radar screening caused by nearby structures and terrain. This plot was obtained by flying an aircraft at constant altitude around the test site, keeping the target at constant range and aspect. A plot of the relative radar reflection cross section of an aircraft is shown in Figure 12. The aircraft was flown in circular pattern of approximately 10 miles diameter at a distance of 10 miles from the test site. As the aircraft maintained a constant 5 degree bank, two halves of the plot are taken from 5 degrees above and 5 degrees below the horizontal plane of the aircraft.

Track Comparaters

Jamming of a tracking radar is effective if the jammer causes the radar position and velocity output data to become so noisy that the data is unsuitable for effective weapons system control or provides no track at all. The method chosen for this evaluation was an instrument that would continuously record position information from a tracking radar being jammed and position data from an unjammed tracking radar, in a form suitable for the Datatron 205 computer. Position information from the two radars was obtained from the range, azimuth and elevation servos of the radar. As all this information is recorded on punched paper tape, the computer is programmed to time correlate all the information to specified time increments. Parallax correction is also programmed into the computer to correlate all the data of the jammed radar to the controlled tracking radar. The correlated data of the jammed radar is then compared to the reference radar and position error calculated.

SUMMARY

The instrumentation developed has demonstrated the feasibility of automatically measuring the performance of jammers and ECM in terms of Self Screening, Area Screening, Deception, and Signal to Noise. Even though the breadboard models suffer from instability and drift, they have proven capable of performing measurements with quantitative precision and repeatability far better than is possible using the human operator. In the course of testing the S/N meter numerous additional applications were discovered such as dynamic measurement of radar cross-section, radar screening angle, and radar antenna pattern.

A new program will use results from these breadboard models to provide improved model JE Meter for the AN/FPS-20 and AN/MSQ-1A radars. Several changes will be incorporated:

1. The Gating Group accuracy will be improved so that sensitivity should increase about 6 db relative to the human operator.

2. The noise sampling system will be changed from the present linear averaging to one of statistical detection. This will render the equipment virtually independent of the statistical distribution of the jammer noise.

3. The number of separate timing gates will be reduced so that the cumulative errors and instabilities have less effect upon overall accuracy.

In conclusion, we feel that this instrumentation is a material contribution in the area of test instrumentation. More than that, in the field of ECM testing where the difference between Engineering tests and Operational tests is often a difference only in the number of items tested, this equipment represents a significant step toward truly engineering type measurement.

REFERENCES

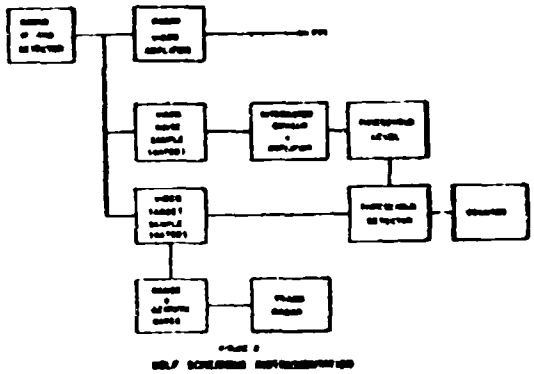
1. [illegible]-TR-58-328 / AD-303268 (ASTIA) — Second Interim Engineering Report Jamming Effectiveness Instrumentation Study 24 January 1958

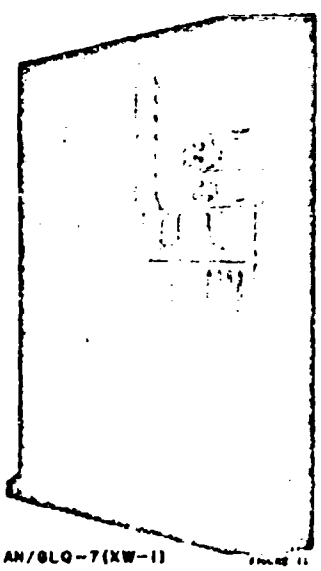
2. RADC-TR-58-381 / AD-303861 (ASTIA) — Third Interim Engineering Report Jamming Effectiveness Instrumentation Study 24 May 1958

3. RADC-TR-59-257 — Fourth Interim Engineering Report Jamming Effectiveness Instrumentation Study 15 June 1959

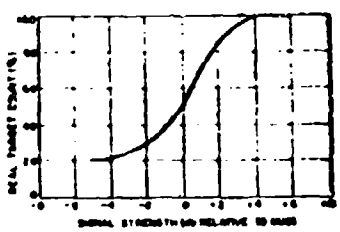
4. RADC-TR-59258 — Fifth Interim Engineering Report Jamming Effectiveness Instrumentation Study 2 July 1959

5. RADC-TR-59-165 — Final Engineering Report (Phase I) Jamming Effectiveness Instrumentation Study 4 September 1959

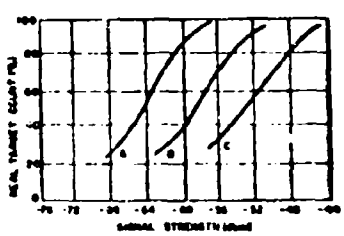




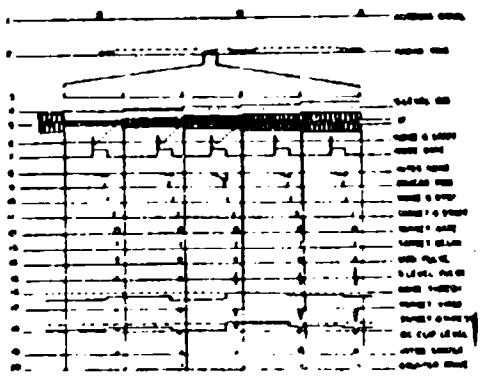
AN/SLQ-7(XW-1)



DETECTION CALIBRATION CURVE



EFFECTS OF NOISE LEVEL ON DETECTION CALIBRATION CURVE



TIMING CHART



SYNTHETIC TARGET SYSTEM, CALIBRATION CURVE

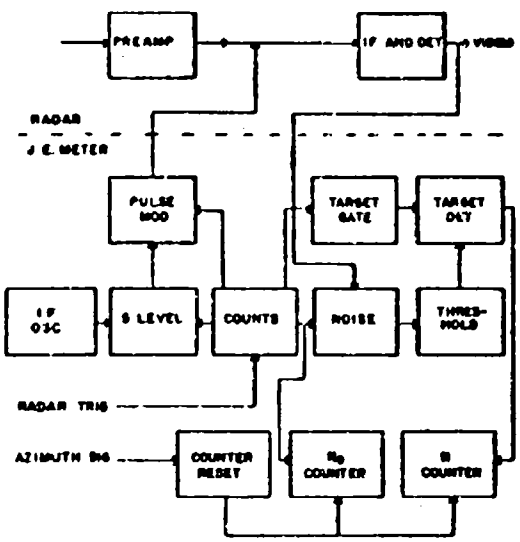
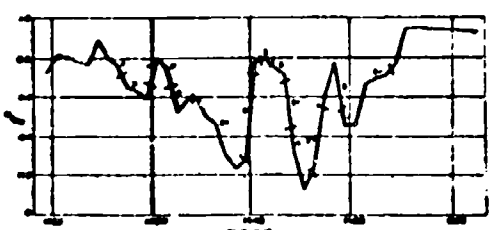


FIG 5
AREA SCREENING BUSSET



AREA SCREENING

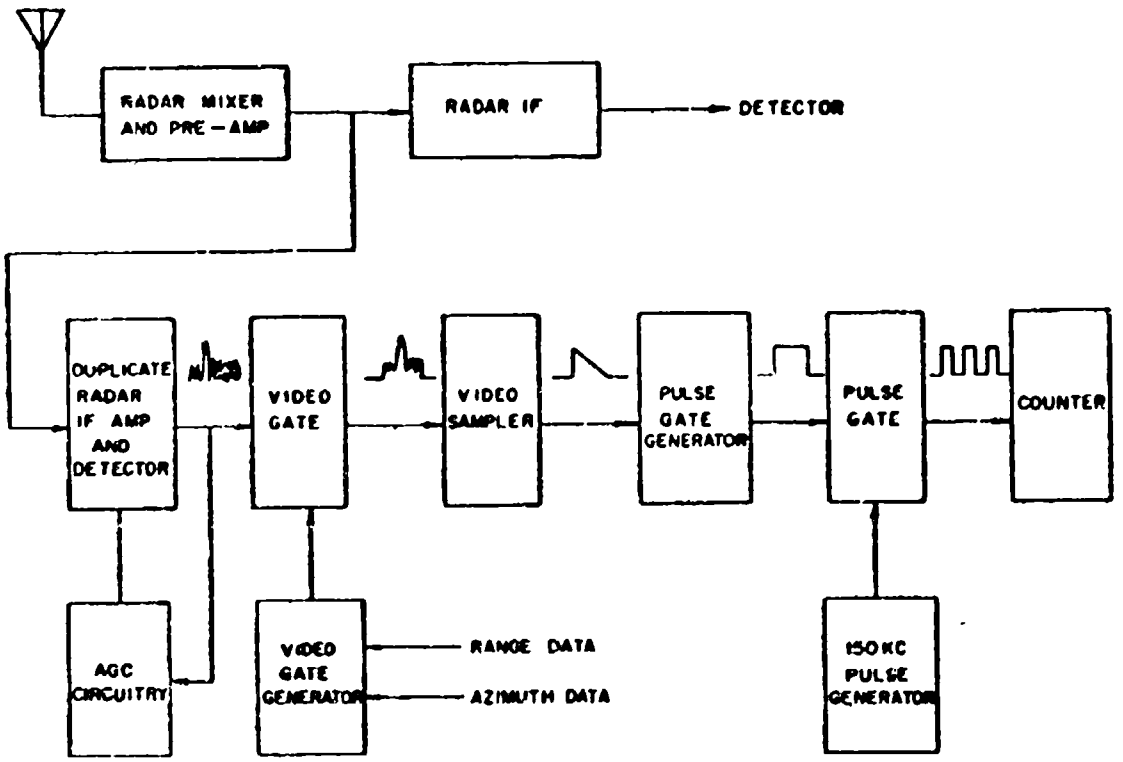


FIGURE 9

SIGNAL TO NOISE INSTRUMENT

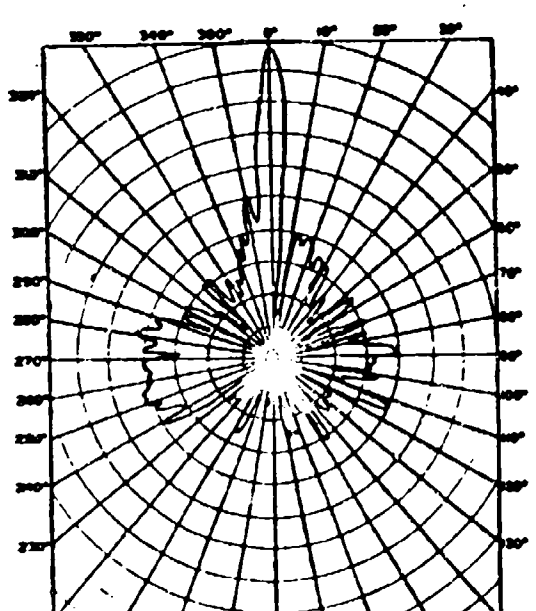


FIGURE 9

HORIZONTAL PATTERN OF RADAR ATENNA AN/TPS-1D

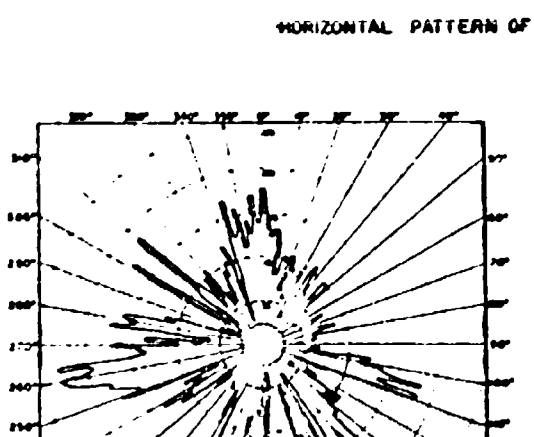


FIGURE 10

GROUND SCREENING

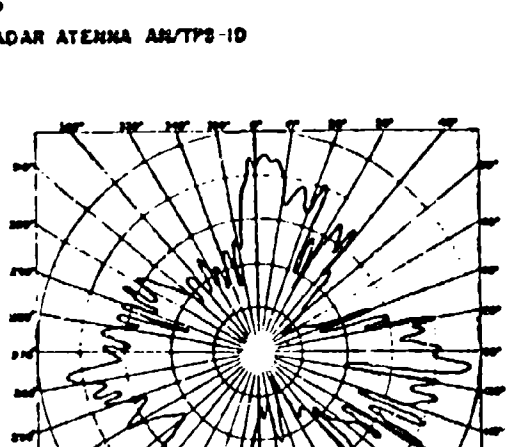


FIGURE 11

RELATIVE RADAR CROSS SECTION OF KC-135 AIR CRAFT

HIGH-SPEED AUTO-DATA SYSTEM FOR BLAST STUDIES

By: Mr. R. D. Jones and Mr. J. B. Smith, Sandia Corporation

Introduction

Recognizing that an automatic data handling system would greatly shorten delays between the collection of raw data and final reduction of these data to a form suitable for analysis, Sandia Laboratory has developed a multichannel, high-speed, automated data system for recording digitally the outputs of the analog transducers used in the study of blast effects from small high explosive charges (about 10,000 pounds of TNT. See Fig. 1). Test phenomena associated with these studies are of short duration (10 to 40 milliseconds).

To satisfy the requirements peculiar to blast-study work, a system was needed that would:

1) Provide automatic data reduction in the field

2) Deliver recorded information, for more detailed analysis, to conventional computers in a form compatible with their input requirements

3) Respond over a frequency range from DC to 5 kc or more, with a rise-time capability of 0.1 millisecond

4) Have an accuracy of 1 per cent of full scale.

The system developed to meet these requirements provides for the sampling of each of 32 analog inputs at a 12.5-kc rate and for the recording of sample amplitudes in the form of a serial, seven-digit, binary number on magnetic tape. This means that the information handling capacity of the recording system must therefore be 2.8 megabits per second. Allowing an additional 400 kilobits for internal synchronization and channel identification, the total recording capacity of the system must be 3.2 megabits per second.

After recording, the magnetic tape is played back at a rate slow enough to permit retranscription of the information to seven-level paper tape. The paper tapes are then read by auxiliary equipment, and the information is decoded and plotted on standard graph paper in a form suitable for "at the site" examination. Alternatively or concurrently the information can be translated to binary decimal excess-3 format suitable for input to an Elecom 125 computer for more detailed analysis.

System Organization

A simplified block schematic of the complete system is shown in Figs. 2 and 3. The recording section (Fig. 2) is straightforward and illustrates the sequence of sampling-coding-recording operations previously described. The playback section (Fig. 3) summarizes the operations required to provide a secondary transcription of the data to punched paper tape.

Functional details, as well as descriptions of the various major components, will be presented in the following sections. They will be grouped according to their relevance to data coding-recording, playback, or processing.

Recording

Commutator -- As shown in Fig. 2, blast-induced transients in analog form, detected by transducers located in the immediate blast environment, are transmitted by wire links to a commutator which assembles samples of the inputs in a time division multiplex to a high-speed analog-to-digital converter.

Individual channel switches, using solid-state devices and printed circuit techniques, are packaged in groups of eight and are activated at a 12.5-kc rate by control pulses from the program control unit. In the interest of maintaining modular flexibility, provision was made for switching eight channel groups at rates of either 6.25 kc or 25 kc through the use of a patch panel. In principle, commutation could be accomplished at still lower rates, thereby increasing the number of channels. However, the particular requirements of the blast studies group were satisfied by the 32-channel, 12.5-kc sampling-rate arrangement.

Forward resistance of the individual switches in the "closed" position is 50 ohms maximum to a common load of 1000 ohms. Back resistance of the switch in the "open"

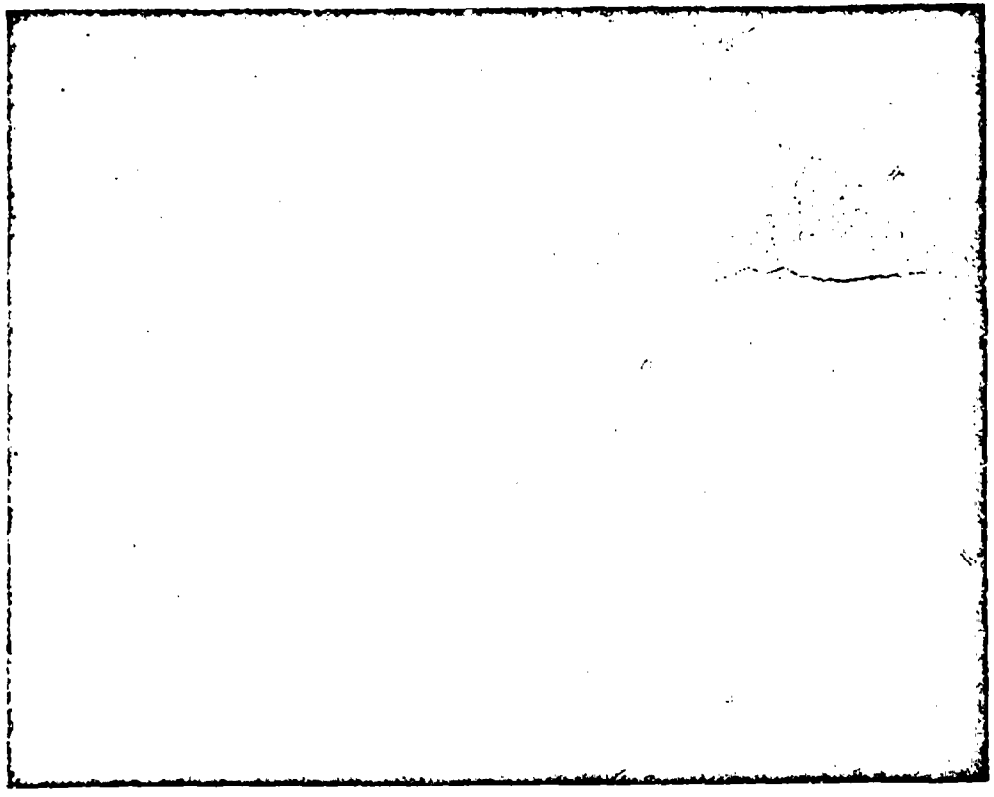
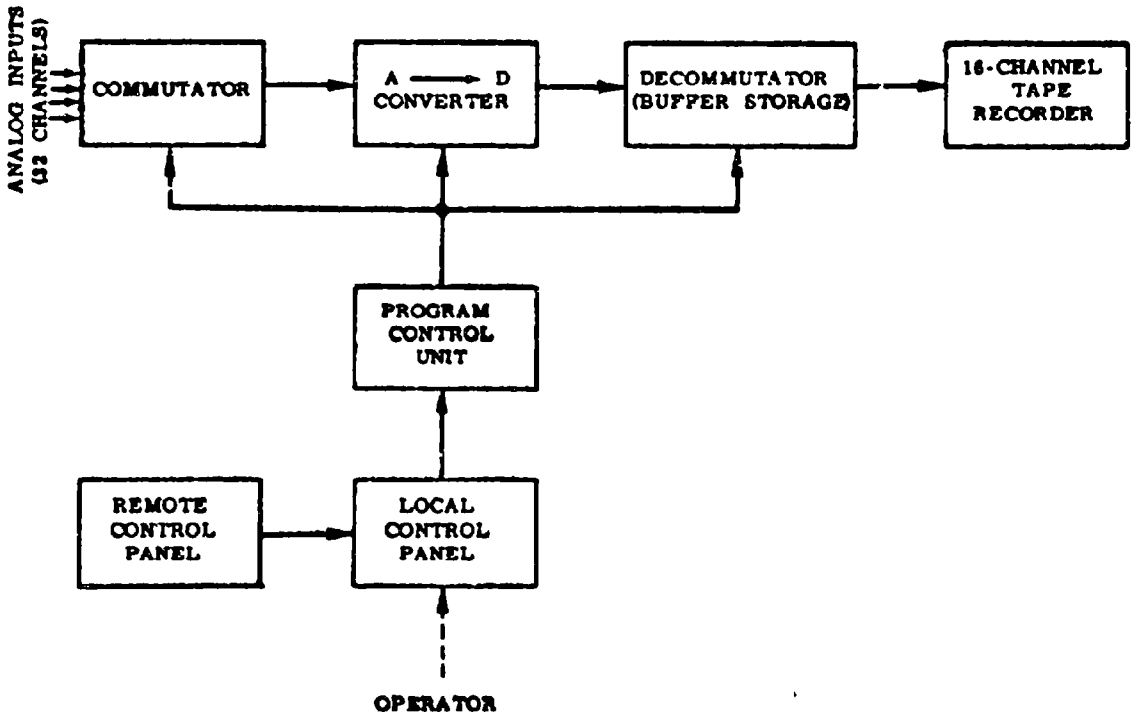
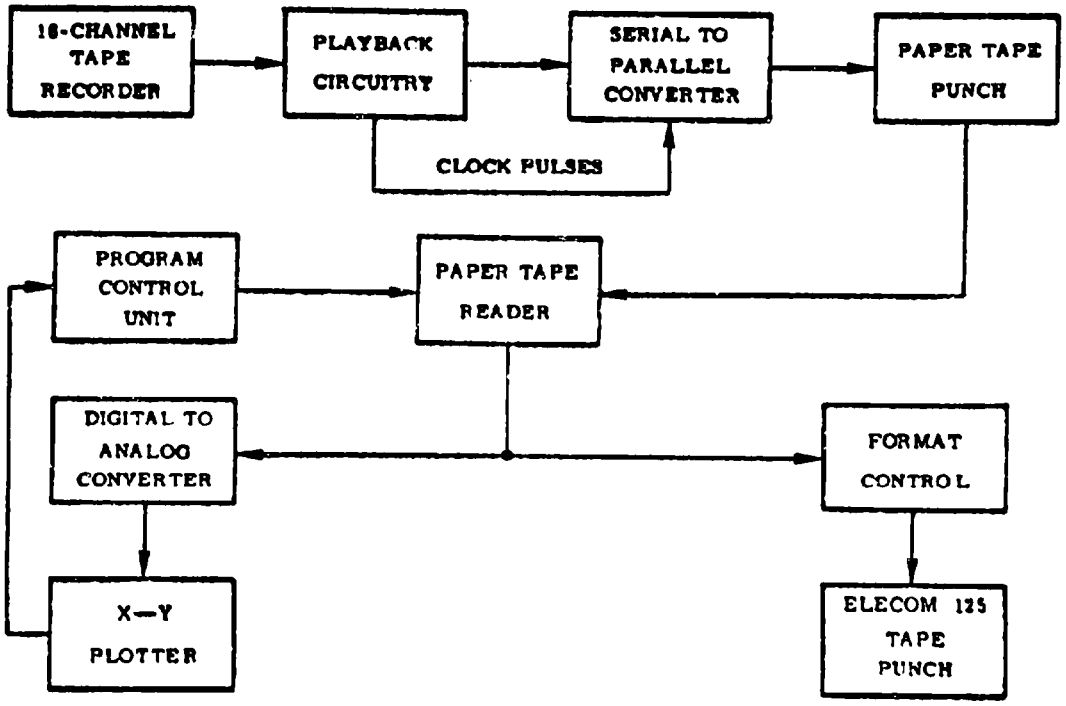


Figure 1



Functional block diagram AUTO-BLAST-DATA-SYSTEM (recording)

Figure 2



Functional block diagram AUTO-BLAST DATA-SYSTEM (Playback)

Figure 3

condition is 4 megohms. Maximum signal input level is 1 volt.

Analog-to-Digital Converter -- In view of the short period (i.e., 2.5 μsec) available for digitizing, the "flashcoding" cathode-ray tube techniques developed by R. W. Sears and associates at the Bell Telephone Laboratories were utilized to obtain analog-to-digital conversion. Tubes of this type have accomplished conversion at rates as high as 10 mc.[1]

Because the converter is the heart of the recording system, it deserves further consideration here. The essential feature of the coding tube is a perforated plate, the holes of which represent an ordered array of the 2^n amplitude levels characterized by an n-digit binary code. A ribbon electron beam is deflected vertically by the output of the commutator.

Separate collector plates are placed behind the code plate. A pulse is produced only if the beam goes through a hole in the plate. Since the ribbon beam covers the full width of the code plate, the code is registered simultaneously on a plurality of digit collectors, one for each digit of the code. It should be noted that the digital outputs are in parallel here and that the code employed is the Gray code.

Decommutator -- The primary function of the decommutator is to serialize the output of the coding tube. Then the seven-digit number characterizing the amplitude of a sample from a given channel is interlaced with the seven-digit number representing the amplitude of a sample from a second channel to form the two-channel serial multiplex depicted in Fig. 4. That is, digital representations of channels 1 and 3 are contained in an 80-μsec frame corresponding to the 12.5-kc sampling rate. (The term "frame" is used here to refer to the particular array of digit pulses associated with a given sample. The term "framing pulse" refers to the timing index or fiducial pulse used to mark the instant of sampling and to order the array.) In this way 32 channels of data are recorded on 16 magnetic tape channels.

It should be noted (Fig. 4) that the channel-3 framing pulse is redundant since channel 3 is sampled 5 μsec after channel 1. The space which would otherwise be assigned to the channel-3 framing pulse is reserved for a playback program-control pulse used as an event mark. This pulse can also be made to activate the paper-tape punch in playback, thereby providing a search-mode operation.

The decommutator provides one other function, that of inserting a recognition signal consisting of all zeros in the frame except for the framing and playback program-control pulses. This unique value is reserved to show that calibration information will follow on the channels.

The first step in the decommutation process is to feed the seven-digit outputs of the analog-to-digital converter to magnetostrictive sonic delay lines which provide delays varying from 10 to 70 μsec in 10-μsec steps. After amplification and regeneration, the digital pulses (delayed for longer periods as their significance decreases) are fed to a solid-state logic complex where they are combined with the previously described program signals to make up the format of Fig. 4.

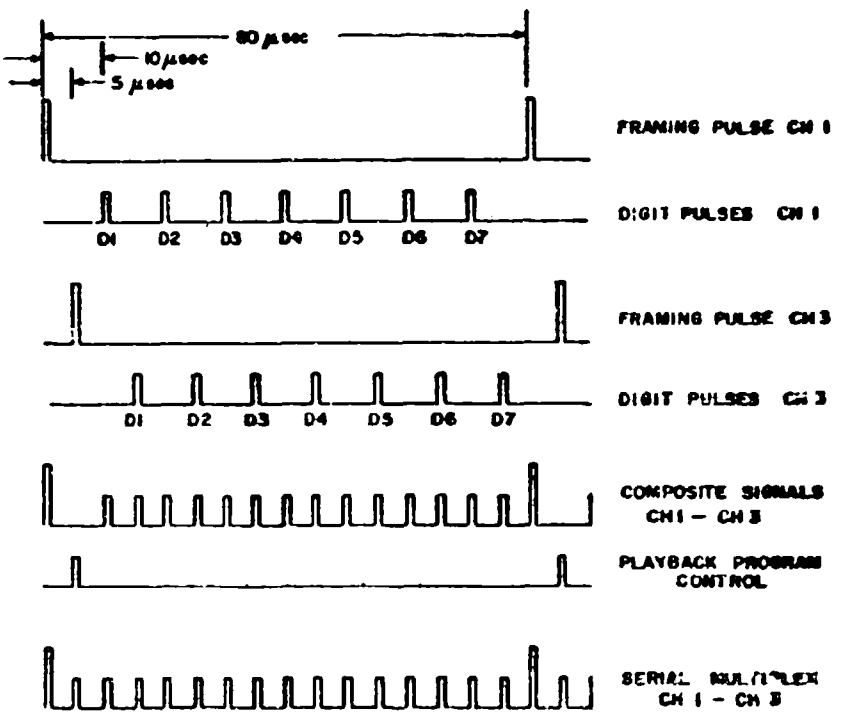
Finally, after power amplification, the 16-channel output of the decommutator is fed to the magnetic tape recorder.

Program Control Unit -- Obviously, the success of the timing and gating operations described above depends upon accurate time alignments of control pulses from the program control unit. These are obtained from a set of 12 basic control pulses which are fed to various units where they are combined in diode matrices to form control pulses for commutation, decommutation, etc. For convenience, a standard printed, diode, plug-in package consisting of eight 3-leg AND gates was adopted for this purpose.

For the most part, individual circuits are conventional. A few of the more unconventional circuits have been described by Meacham and Peterson.[2]

Recorder

Serial-type magnetic tape recording was adopted in preference to parallel tape recording. The playback circuitry required by the serial type is not as extensive but is more functionally complex than the parallel type, since interlaced frame and data pulses must be separated electronically. The complexity arises in reclocking the pulses at the playback rate. Although this electronic discrimination is



WORD FORMAT
32-CHANNEL BLAST DATA RECORDER

Figure 4

avoided in parallel recording in which data and framing pulses are recorded on separate magnetic channels, pulse amplification and regeneration equipment must there be duplicated.

An additional justification for the adoption of serial recording stems from a consideration of the mechanical problems associated with parallel recording. In particular, registration difficulties from variable differential tape stretching at high recording speeds were anticipated. These difficulties become even more acute in field use where tape handling and storage facilities are far from ideal and where both tape and recorder will be subject to extremes of temperature, humidity, and dust (see Fig. 1).

The tape transport employed was developed for this particular application by Clevite-Brush. Sixteen channels are recorded on 3/4-inch Mylar-base magnetic tape fed from a basket providing storage for 800 feet. The recording speed is 350 inches per second with a recording pulse density of 570 pulses per inch, which corresponds to the 200-kc pulse recording rate.

Once serial-type recording was adopted, it was decided for simplicity to differentiate the framing and data pulses by recording at two amplitude levels (see Fig. 4). Since the framing pulse appears only once for each frame, it was assigned maximum amplitude. The limitation on the amplitude of the framing pulse is fixed by saturation of the tape magnetization. Of course, while the code pulses must have lower amplitudes to distinguish them from the framing pulse, their amplitudes must be high enough to maintain a favorable signal-to-noise ratio on the comparatively slow-speed playback.

Playback

The primary function of the playback system is to transfer the data recorded on the primary storage medium (magnetic tape) to permanent storage in the form of punch paper tape (see Fig. 3). To do this, the magnetic tape is examined channel by channel; in a 32-channel set-up, 32 magnetic tape passes are required to punch 32 paper tapes, one for each of the original analog input channels.

The first step in this process is amplification and regeneration of the recorded pulses. The playback circuitry provides pulses of constant width and amplitude and also separates out the framing pulse on the basis of amplitude discrimination. The framing pulse drives a clock-pulse generator which produces two sets of eight clock pulses corresponding to the positions of digits indicated on Fig. 4 for channels 1 and 3. A "fly-wheel" effect is included in this circuit so that loss of an occasional framing pulse will not cause loss of synchronization.

Various magnetic-tape playback speeds were tried, and the results indicated that 1 inch per second was the minimum acceptable playback rate for a signal-to-noise ratio great enough to ensure reliable synchronization of the recording and playback circuitry. With a recording speed of 350 inches per second, the minimum acceptable playback rate is approximately 35 frames per second.

The final step in recovery of playback data consists in making a serial-to-parallel conversion. Regenerated data pulses corresponding to either channel 1 or 3 can be gated into a shift register composed of a linear array of two-cores-per-bit magnetic storage elements. In response to shift signals from the clock pulse generator, the data pulses move along the array until the register is completely filled. At this point, the data in the shift register are transferred to a seven-level Teletype tape punch. Temporary parallel storage and interlocking electronics are required to synchronize perforation with the playback framing rate. Data characterizing amplitudes of successive samples of each of the analog inputs have now been transcribed to seven-level perforated tape and are available for processing.

Data Processing -- The computing-plotting facility developed for preliminary data evaluation consists of three principal units: a paper tape reader, a digital-to-analog converter, and an X-Y plotter. A fourth unit, an asynchronous program control, coordinates operations of the other units mentioned. Digital data from the punched tape are translated by a digital-to-analog converter into analog signals for automatic point-by-point plotting by the X-Y plotter.[3] An asynchronous program control was devised so that the plotter would receive information from the reader only as fast as it could be plotted.

In general, successive plot transits are small enough to permit plotting at rates from 2 to 4 points per second. In cases in which more elaborate mathematical operations on the recorded data are desired, the outputs of the paper tape reader are fed to a converter which provides a binary decimal excess-3 retranscription of the same data in a format suitable for input to the Elecom 125 computer.

Installation

The complete automatic data system is installed in an underground shelter at Sandia Laboratory's Coyote Canyon Test Field. Regulated power supplies are mounted in the right end bay shown in Fig. 5. Playback equipment is located immediately to the left of the power supplies. The next bay contains the 16-channel magnetic tape transport (common to both recording and playback operation). Directly to the left of the tape basket are the flash coder (analog-to-digital converter) and the sonic delay lines. These chassis have been turned in their racks for the illustrative purpose. The commutator and high-speed program control unit are located above the delay lines.

Not shown is the control console which contains the teletype tape punch, reader, X-Y plotting equipment, and control panel.

Operation

In practice, the control console is supervised by a remotely located master program timer. In response to control signals from this unit, the recording equipment is automatically activated. Immediately before detonation of the explosive charge, the high-speed forward traverse of the tape transport is initiated. After zero-time signals and calibration information for individual channels are recorded, all input channels are cleared to record outputs of the analog transducers employed to detect the blast effects. After the shot, the high-speed forward traverse of the magnetic tape is halted, and control of the system is returned to the local operator by the master program timer.

At the discretion of the local operator, the playback circuits are then activated and the magnetic tape played back. Recorded data, including zero time marks and calibration information, are retranscribed on paper tape. The paper tapes are then ready for processing.

Conclusion

Despite the high ratio of recording to playback speeds and the use of conventional recording-playback techniques, synchronization can be maintained between recording and playback equipment even under the environmental extremes encountered in blast study work.

The authors originally hoped that enough reliability could be designed into the system so that it could be maintained and operated by untrained personnel. Unfortunately this aim was not realized; one of them was always required to "mother" the system. It should be emphasized, however, that no attempt has been made to redesign the system for greater reliability; rather, any design expedient that seemed to promise earlier field operation was adopted.

As is often the case in blast study work, the equipment described here was hurriedly assembled for a particular test series of blast diffraction experiments and was utilized for the evaluation of test instrumentation. The details of the system are reported here, not only for their historial interest, but also because they illustrate the application of digital instrumentation techniques to a rather unusual area of endeavor.

Because of sentimental attachment to the system by the authors and their supervision, the device was not consigned to salvage--the usual practice--but has rather been retained almost intact on a standby basis. It is now used as a training aid to instruct new personnel in the principles of digital instrumentation.

Acknowledgements

The authors are grateful for the cooperation of the following Bell Telephone Laboratories personnel: Mr. R. W. Sears, for arranging the loan of a seven-digit flash coding tube, and Messrs. R. K. Potter and K. L. Carbrey, who supplied designs for associated circuitry.

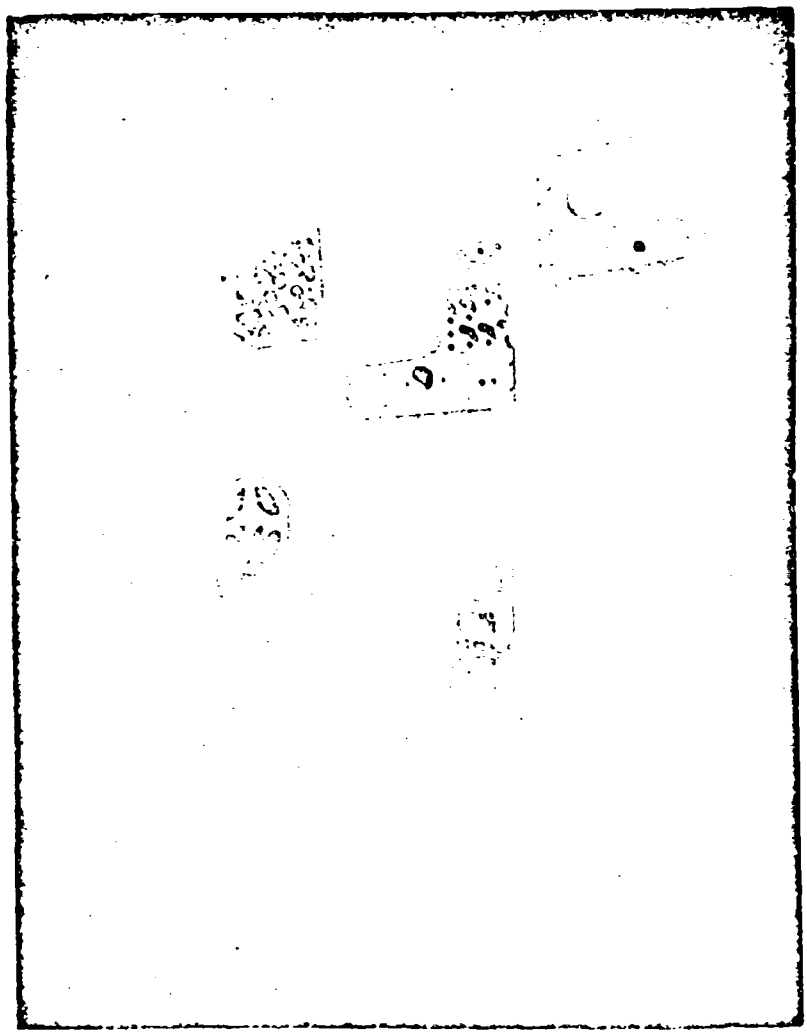


Figure 5

This work was supported by AEC Contract No. AT-(29-1)-789.

Bibliography

1. Goodall, W. M., "Television by Pulse Code Modulation", Bell System Technical Journal 30, 33-49, January 1951

2. Meacham, L. A. and Peterson, E., "An Experimental Multichannel Pulse Code Modulation System of Toll Quality", Bell System Technical Journal 27 (1) 1-43, January 1948

3. Robnett, A. V. and Vulgan, E. J., "Small-Scale Automatic Field-Data Reduction System," Sandia Corporation AECU-3123, October 3, 1955 - Office of Technical Services, Department of Commerce, Washington 25, D. C.

DATA HANDLING FOR A RESEARCH AND DEVELOPMENTAL STATIC TEST STAND

by

Thomas Wong and Robert L. Thomason
U. S. Naval Ordnance Test Station

ABSTRACT

This paper describes a data acquisition system for a static test complex being designed and built at the U. S. Naval Ordnance Test Station, China Lake, California. The system can measure phenomena from rocket motors having an average thrust of 1,000,000 pounds and a peak thrust of 10,000,000 pounds. The data measurement capacity is 200 channels.

The test complex is designed to be a research and developmental tool rather than a production-testing facility, with emphasis on flexibility in accomodating new test programs together with ease and minimum cost of operation. As such a tool it is used to test new components and methods and to advance the state of the art.

Discussion is concentrated on the technique of data handling and measurements, real time digitization, automatic calibration and stored program data control.

Included will be a description of a new sampling technique for commutating multiple data channels. The sampling format is programmable and can be altered during a test according to existing conditions.

INTRODUCTION

A new test facility with automatic data handling capabilities is now being built at the U. S. Naval Ordnance Test Station for static tests of large propulsion systems capable of thrusts of a million pounds or more. Preliminary layout and system designs were begun one and one-half years ago, and the construction of major segments has been completed.

The design of the test facility represents a team effort on the part of data-handling groups, propulsion engineers, test and evaluation people, and design engineers who have tried to approach as closely as possible the ideal system--from missile handling to final printed data.

The plant is designed not to be a routine or production testing facility but rather a research and developmental tool for extracting a maximum amount of meaningful data from a minimum amount of test effort. In line with this aim, flexibility is one of the key items being incorporated in the system, so that new test programs can be conducted with minimum restrictions, downtime, and cost.

TEST STANDS FOR MOTOR TESTING

The facility will have two air-conditioned bays, each containing a test stand. The bays can easily accommodate motors six feet in diameter and 50 feet in length; each test stand can handle average thrusts of one million pounds and peak thrusts of ten million pounds. New approaches to force measurements and structural design were embodied in the construction of the test stands. Loading conditions that may affect the stands were simulated on the Station's REAC analog computer to check theoretical factors. Features for six components of force and instantaneous weight measurements were included.

The first stand, which has been completed and is now in operation, is of the horizontal type. Construction of the second stand will be started this summer. This stand will have a movable thrust block to allow for both horizontal and vertical capabilities. Motors to be tested are of the solid-propellant type. Handling of liquid hybrid propellants is included in future expansion plans.

DATA-HANDLING PHILOSOPHY

The method of data handling depends upon the test purpose. For data designed to measure the parameters of a motor and to provide for performance evaluation, high accuracy is required and is typically limited to relatively low frequencies. The primary data for such performance evaluations is the thrust-vs-time measurements, total impulse and specific impulse being the two most important parameters. The various types of data are processed to give these two parameters the maximum degree of accuracy and confidence attainable within the present state of the art.

Thus the basic data requirements are quite simple (thrust and specific impulse) but, in order to provide the history for design considerations and evaluation of components and materials of a motor and also to provide independent measures of the various internal ballistics parameters, other data are required. Measures of pressure, temperature, stress, acceleration, weight or mass change, vs time are needed to insure that the design goals can be reproduced.

All of the data acquired and processed is used to calculate the basic parameters and to evaluate the motor under different conditions. Since the data acquired from test to test forms a

history of expected performance and malfunctions in general are a deviation from the normal, such of the data acquired is used only in the event of a malfunction. Therefore the data acquisition and processing system must provide a means of scanning vast amounts of data and sorting out only that part which is pertinent to the parameter being studied.

OVERALL SYSTEM DESCRIPTION

Figure 1 is a block diagram of the overall system. Over 200 transducer channels feed from the motor in the test stand into the patching network in the terminal building next to the bay. From there they connect through land lines in an 800 foot long underground tunnel into another patching network in the instrumentation building. The lines have 100% shielding to eliminate noise and crosstalk. A similar arrangement will also feed lines from the second test bay into the instrumentation building. The data channels are divided in different modes of recording depending on the frequencies carried on any particular channel and the accuracies to be achieved. Quick-look recording equipment provides a direct indication of a selected number of channels during the test. Wideband signals are recorded in FM record mode. Medium and low frequencies are recorded as a composite of standard IRIG bands in direct-record mode. The more critical functions are processed through a digital data processor in real time for the determination of high-accuracy data. The digital data processor is restricted to dealing with low frequencies because of the limited sampling rate achievable. In the event of malfunction of the digital recording the data is not lost; it can be recovered from the medium frequency data by playback of the analog recording into the digital data processor. An IBM 709 computer is used to reduce the data.

All the equipment (except the IBM 709 computer) is located in the instrumentation building to allow last-minute accessibility, elimination of duplicate terminal equipment for the two bays, and to insure its survival in the event of test-bay destruction due to motor explosion.

SYSTEM FEATURES

Transducers

The ultimate limit on the measurements lies in the precision of the physical sensing elements, namely the transducers. To insure the highest

DATA ACQUISITION AND STORAGE SCHEMATIC

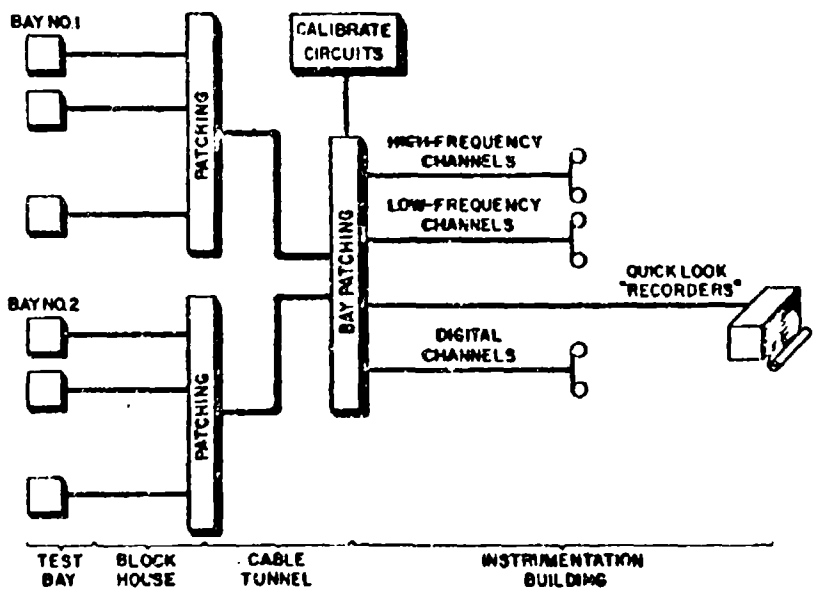


Fig. 1. Data acquisition and storage schematic

possible precision a modern calibration laboratory is included in the facility. Laboratory equipment includes shock tubes for dynamic calibration and such standard items for static calibration as dead-weight testers.

In the test area, for final system readiness determination, an 11-point voltage calibration is made automatically to check out the equipment in the instrumentation building. A delta-R calibration is also made to check out the overall system, including land lines and the transducer signal outputs. This delta-R calibration is part of an 8-wire transducer hookup system. Other features of the 8-wire system include zero adjust and bias offset. The zero adjust takes care of drift; bias offset allows selection of the dynamic range for a particular application.

Because of the 800-foot length of the lines from the test stand to the instrumentation building, a direct current method of exciting the transducers is used. The DC method avoids some of the difficulties involved in using an AC carrier system, such as those due to capacitance in the transmission line and the limited accuracy of AC equipment.

Patching

Because there are two test bays, the data-acquisition system is set up so that the instrumentation equipment for the 200 channels can be switched from one bay to another with minimum downtime. Cables from both test bays are brought up to the instrumentation building and are connected to taper pin blocks. From this, they are connected to a master patchboard. By changing the patchboard, the common instrumentation equipment is automatically switched. No downtime is necessary nor is there any interference with the ordnance work crew. In fact, two test firings of large motors within minutes of each other are possible and practical. That is to say, once the electronic equipment is set up and calibrated, it makes no difference which bay the data is coming from. Subpatching is used to segregate major equipment, such as tape recorders, amplifiers, calibration equipment and the digital data processor. This feature allows independent maintenance and availability of the equipment.

Signal Conditioning

Signal conditioning equipment consists mainly of instrumentation-type DC amplifiers and filters. The selection of the amplifiers is made for high common mode rejection or wideband signal requirements. The choice of the filters is based not only on the intelligence frequency, but also on the sampling rate.

Quick-Look Recording

Forty-two channels are available for a direct write quick-look presentation of frequencies up to 5,000 cycles. The patching allows a quick look of any other desired channels on playback after a firing.

Analog Tape Recorders

Regardless of the mode of presentation or conversion, all data channels are recorded on magnetic tape. Figure 2 shows a typical analog tape recording channel. For low frequency recording IRIG-standard multiplexing techniques are used. For high frequency data wideband single-channel recording is used. The analog recordings can be played back into the digital data processor. For digitizing high-frequency data, time-base expansion techniques are applied.

Digital Data Processor

The digital data processor digitizes data in real time to bypass inaccuracies caused by modulation and demodulation operations, and converts the data into the proper format for entry into the IBM 709. It is also used for monitoring and control purposes.

The digital data processor is a completely transistorized unit. An overall block diagram is shown in Figure 3. Some of its functional features will now be described.

1. The Commutator. The commutator can select up to 128 channels. Channel selection is made by stored-program command. Thus sampling rate and sampling format can be varied to suit test requirements. Samples are taken at a rate of 12,500 per second.

2. The Sample and Hold Unit. As shown in Fig. 3, a sample and hold unit is employed. This unit stores the sampled input voltage during analog-to-digital conversion time to minimize errors.

3. Analog-to-Digital Converter. Data words are converted 12 bits at a time and accumulated into a 24-bit shift register.

4. Ferrite Core Memory. The 24-bit words are fed in parallel into a 2048-word ferrite core memory.

5. In-Output Buffer. The in-output buffer is used to store each 24-bit word and then shift it in 36 us intervals, 6 bits at a time, in the tape-write circuitry for the digital tape machine.

6. Tape Machine. Tape-machine features include a tape speed of 150 inches per second and a bit density of 200 bits per inch.

7. Flexowriter. A flexowriter is used as an in-and output-device for system checkout.

8. Card Reader. A photo-electric card reader is used along with a tape file to provide a quick program read-in for the digital data processor.

9. Central Control. The central control unit initiates and controls the proper action for the various parts of the digital data processor.

10. Manual Control. The manual control unit provides a single instruction mode, without interfering with the regular program.

11. Memory Control. The memory-control unit provides the proper memory access according to a well-defined priority system.

For the future, data-handling equipment for microwave transmission is being planned to connect the test facility directly to the IBM 709 for real-time monitoring and computing during a test firing. Also the digital data processor is being considered as support for other evaluation and plant process data handling problems.

CONCLUSION

A new test facility with automatic data-handling equipment is being built to accomodate the testing, development, and evaluation of large rocket motors. Normal thrusts of one million pounds can easily be handled. The instrumentation will have a data-handling capability of 200 channels with real-time digitization of 128 channels. This data system incorporates the latest techniques and equipment. Incremental construction techniques are being used to gain valuable experience before the complete system is finalized.

The system design philosophy has been toward flexibility without loss in accuracy or reliability. The digital processing scheme is complex, compared to conventional data acquisition and handling systems, but it is versatile, reliable, self contained, easy to maintain, and essentially free from human setup errors.

The overall system is thus designed to meet its mission as an experimental and developmental tool while advancing the data acquisition state of the art.

ACKNOWLEDGMENT

Although this project has been a team effort on the part of too great a number of people to single out individual names, special recognition should be made of the contributions of the Propulsion Development and Test Departments of the U. S. Naval Ordnance Test Station.

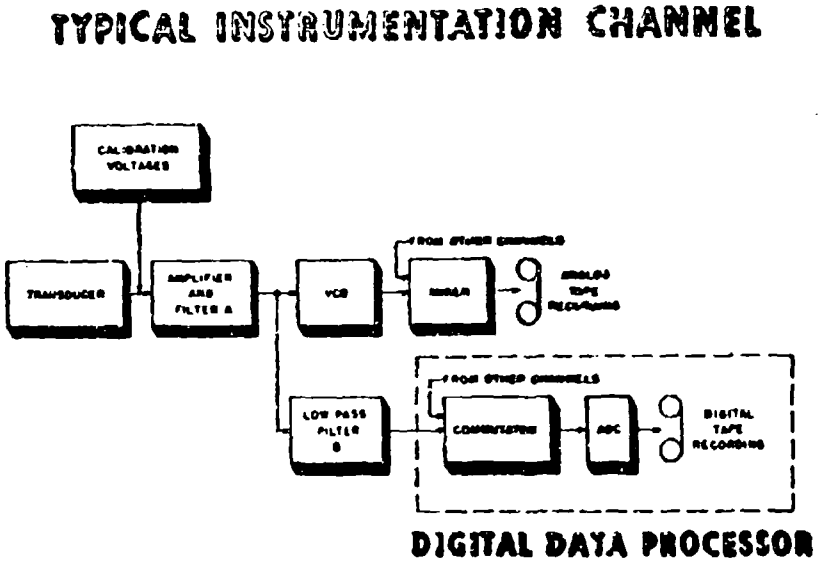


Fig. 2. Typical instrumentation channel.

DIGITAL DATA PROCESSOR

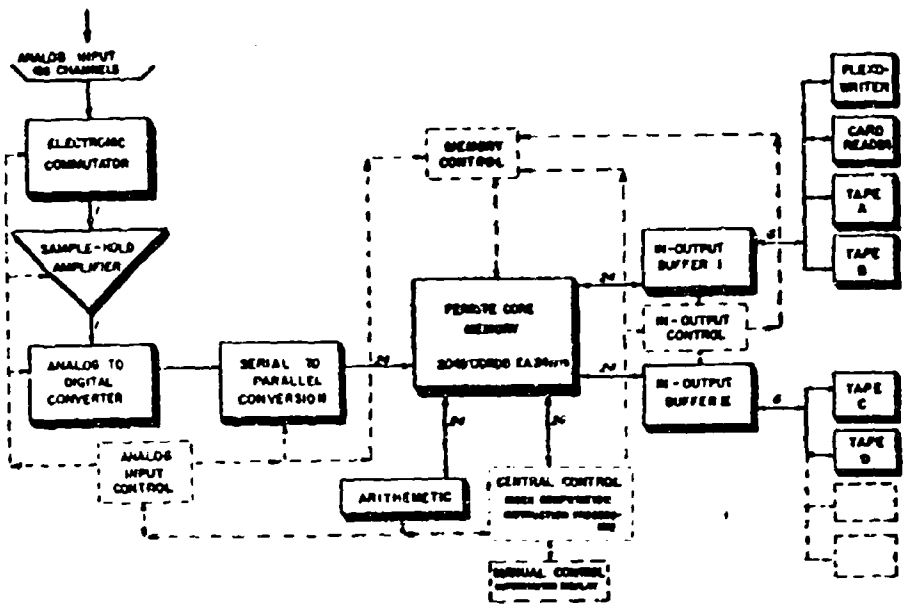


Fig. 3. Digital data processor.

THE HANDLING OF UHF DOPPLER DATA

By: Mr. D. H. Parks, RCA Service Company, Missile Test Project, Patrick Air Force Base, Florida

ABSTRACT

The methods used in processing UDOP (UHF Doppler) data at the Atlantic Missile Range are outlined.

A. INTRODUCTION

UDOP is a continuous wave electronic tracking system which utilizes the integrated doppler effect to obtain range sum or range difference measurements. The system was implemented at the Atlantic Missile Range by ABMA as an improved UHF version of the earlier DOVAP system. The station geometry consists of two independent arrays. One array, consisting of seven receivers and a transmitter within a twenty mile radius of Cape Canaveral, furnishes data from lift-off through midcourse. Another array, consisting of four receivers and a transmitter within a twenty mile radius of Carter Cay, furnishes terminal trajectory data. The doppler data generated at each receiver are transmitted by cable or by an RF link to a recording site centrally located within each array. At the present time no digitization of UDOP data takes place during a flight, all data reduction being undertaken after a test from magnetic tape recordings.

B. DIGITIZATION OF UDOP ANALOGUE DATA

The limiting factor in the development of UDOP data processing techniques has been the wide range of doppler frequencies encountered. The use of a transmitted reference frequency in the neighborhood of 450 megacycles with a return frequency in the neighborhood of 900 megacycles has resulted in doppler frequencies greater than 25 kilocycles at an IRBM burnout. Such frequencies are far beyond the capabilities of the opto-electro-mechanical devices employed in the DOVAP reduction. In order to obtain velocity components with a random error of less than one foot per second throughout the burning period and beyond, it is necessary to measure each total doppler count with a random error of less than 0.1 cycle. These accuracies have been attained at AMR with the use of DARE (Doppler Automatic Reduction Equipment).

DARE was designed to convert analogue data of cyclic form (doppler cycles, square waves, sawtooth waves, etc.) into digitized accumulated cycle counts. In order to accomplish this, DARE performs the following operations:

(a) It counts the total number of positive crossovers (N_c) from the beginning of a data run.

(b) It reads 100-PPS coded timing, digitizing the time (t) at each timing pulse.

(c) It counts the number of 20-microsecond pulses (n) occurring between each positive crossover. These pulses are generated by a 50 KC clock. This number is stored while the pulses in the following cycle are being counted.

(d) It reads out t, N_c, n and the number of 20-microsecond pulses (Δn) occurring between the time of the last positive crossover and the readout time. The total cycle count may be computed from the above information as follows:

B.1 $N_t = N_c - N_o + (\Delta n/n)$

where N_o is the initial corrected count.

The equipment is capable of up to fifty readouts per second. Greater precision in the cycle counting may be attained at high frequencies by playing the analogue data through DARE at half or quarter speed. The slower playback speeds reduce the effective time between pulses from twenty microseconds down to ten and five microseconds, thus increasing the number of pulses within each cycle. The accuracy of equation (B.1) can also be improved by computing the period of the previous cycle from the cycle count over the previous time interval, by correcting for equipment delay, and by modifying Δn by a function of the difference between the counted cycle length and the computed cycle length. A most valuable component of DARE is its tracking filter, or tracking oscillator. The oscillator is phase-locked with the basic frequency on the doppler record. It remains phase-locked as long as the basic frequency exists among the noise on the record. If the doppler frequency is not changing rapidly, it is possible to track through short dropouts, since the tracking oscillator tends to remain at the same frequency that existed before the

ereport. Frequencies as low as ten cycles per second may be tracked. Because the tracking oscillator output is always clean, the possibility of obtaining false counts due to noise spikes on the raw data is eliminated.

C. PROCESSING OF THE DIGITIZED DATA

After being digitized, each station's cycle count is submitted to the 709 Computer for differencing and collating. First, second and third differences are computed, and data from up to fifteen stations are collated on one tape. The differences are scanned for possible digitizing errors before the collated data are submitted to the computer for a position run. If the cycle count differences indicate errors in the digitization, these may be corrected either by following the trend of the differences on either side of the questionable area or by fitting a polynomial to valid first differences in the neighborhood of the erroneous data. The total cycle count correction should be an integral multiple of half cycles. The degree and spread of the chosen polynomial is a function of the forces acting on the missile. A higher degree polynomial is required during the burning period than during ballistic flight. Noisy areas as long as fifteen seconds in duration can be reduced satisfactorily if intermittent valid first differences are available.

At the present time, no satisfactory methods exist for the electronic digitization of doppler data during the first few seconds of flight. These early cycles must be read manually from an oscillograph recording. On every UDOP trajectory reduced so far, it has been possible to begin the reduction at lift-off, computing the initial constants from survey data. However, if the data quality has been poor during the early portion of a flight, theodolite positions translated to the UDOP tracking point have been used to "tie in" the trajectory and improve the subsequent data. Normally, the reduction is begun at launch using data from sites near the launching area. Due to limitations in the present recording equipment, data from the outlying sites are not brought into the solution until their doppler frequencies reach about one kilocycle.

The extent of coverage is dependent upon several factors, the design of the missile's transponder and antenna being the most important.

After all errors in the digitizing process have been determined, the collated cycle counts are submitted to the computer with the corresponding station constants and correction data. Certain limits are placed upon the size of the residuals and the number of iterations to be allowed in order to insure that no computer time is wasted on erroneous computations.

D. EVALUATION OF THE POSITION DATA

After the position run is completed, it is evaluated in several ways. First, the results are compared against check points which have been computed prior to the run either by hand or on a short pass through the computer. The results are also compared against data from other systems. The residuals (observed minus computed measurements) are checked thoroughly. If the residuals are large or show sudden jumps, the reasons are investigated. The computed trajectory is also studied for discontinuities. If anomalies are found which are due to errors in the input data, another computer run is made with further corrections. Five types of evaluations are issued with the published data. One is a general statement on data coverage and quality as determined by observing the raw data on an oscilloscope and by observing the cycle count differences. The second is an estimate of the random errors in the cycle count data as computed by a variate differencing method. The third is an estimate of the internal consistency of the data as indicated by the residuals. The fourth and fifth are the effects of the estimated random and systematic errors on the position data as indicated by the variance-covariance matrix.

E. THE DOPPLER EFFECT AS APPLIED TO UDOP

Let

f = the transmitted reference frequency

r_{xt} = the radial distance from the transmitting antenna to the transponder antenna at time t

r_{it} = the radial distance from the ith receiving antenna to the transponder antenna at time t

c_t = the ambient velocity of propagation at the transponder antenna at time t

Then, ignoring the relativistic time dilation, which can be shown to be less than the random error of measurement at velocities up to escape velocity, we have c_t/f as the distance between wave fronts traveling past the transponder antenna. However, the number of wave fronts detected by the transponder will be decreased by the factor $(\tilde{r}_{xt} f/c_t)$ due to the

transponder's movement through the wave fronts. We thus have

E.1 $f - \dot{r}_{xt}\frac{f}{c_t} = f\left(\frac{c_t - \dot{r}_{xt}}{c_t}\right)$

as the frequency received at the transponder. This frequency is doubled in the transponder. The period T_t of the retransmitted signal will then be as follows:

E.2 $T_t = \frac{1}{2f\left(\frac{c_t - \dot{r}_{xt}}{c_t}\right)}$

The wavelength of the retransmitted signal will be elongated in the direction of the receiver by the factor $\dot{r}_{it} T_t$. The new wavelength λ'_{it} will thus be as follows:

E.3 $\lambda'_{it} = T_t\ (c_t + \dot{r}_{it})$

and the retransmitted frequency f'_{it} will be

E.4 $c_t/\lambda'_{it} = 2f\left(\frac{c_t - \dot{r}_{xt}}{c_t + \dot{r}_{it}}\right)$

This frequency f'_{it} is received at the ground and mixed with 2f, giving the doppler frequency $\dot{N}_{it}$

E.5 $\dot{N}_{it} = 2f - f'_{it} = 2f\left(1 - \frac{c_t - \dot{r}_{xt}}{c_t + \dot{r}_{it}}\right) = 2f\left(\frac{\dot{r}_{xt} + \dot{r}_{it}}{c_t + \dot{r}_{it}}\right)$

The rate of change of the range sum, $\dot{u}_{it}$ is now given by

E.6 $\dot{u}_{it} = \dot{r}_{xt} + \dot{r}_{it} = \left(\frac{c_t + \dot{r}_{it}}{2f}\right)\dot{N}_{it} \approx \left(\frac{c_t}{2f}\right)\dot{N}_{it}$

The above derivation considers only the effects of radial movement away from the stations. There is also a drift in phase due to movement norm l to the range vectors as described in reference (2). The error factor $\dot{r}_{it}\ \dot{N}_{it}/2f$ can be largely corrected if one of the receivers lies near the transmitter. At great distances, the time between the retransmission and the reception on the ground should also be considered during the integration of the doppler effect.

F. AN OUTLINE OF UDOP MATHEMATICAL PROCEDURES

The equations of condition for the UDOP system take the form of the general equations of condition for most continuous wave systems. By rearranging station locations, eliminating various parameters and differencing one equation against another one can arrive at the particular equations of AZUSA, COTAR, SECOR, RAYMIST, SHODOP, TRIDOP and many other C/W systems. The equations of condition (F.3) that follow are modifications of equation (24) of reference (2). The need for computing the average index of refraction is eliminated and the equations are placed in a form common to other instrumentation systems. The measured quantity (range sum or range difference) is placed on one side of the equation of condition and its functional representation is placed on the other side. This simplifies and shortens succeeding operations. The refraction corrections are those which would be applied to a range which has been computed with a vacuum velocity of propagation. Vector notation is used for compactness.

Let

$\vec{P}_{mt}$ (x,y,z) = the position of the missile at time t

$\vec{P}_x$ (x,y,z) = the position of the transmitter

$\vec{P}_i$ (x,y,z) = the position of the ith receiver

$\vec{P}_{mc}$ (x,y,z) = the initial position of the missile

P_o (x,y,z) = the approximate position of the missile

C = the vacuum velocity of propagation

f = the transmitted reference frequency

N_{it} = the number of doppler cycles generated at the ith station from the initial time until time t

Δ_{it}, Δ_{xt} = refraction and other range corrections

It is necessary first to evaluate the doubled vacuum wavelength ($\lambda_o = C/2f$).

Then the initial constants, the total phase along the transmission paths at the initial time, are evaluated. Let N_{ic} represent the

number of wave fronts along the transmission path at the doubled reference frequency. Then

$$F.1 \quad \lambda_o N_{ic} + \Delta_{xc} + \Delta_{ic} = \left|\vec{P}_{mc} - \vec{P}_x\right| + \left|\vec{P}_{mc} - \vec{P}_i\right|$$

and

$$F.2 \quad N_{ic} = \left(\left|\vec{P}_{mc} - \vec{P}_x\right| + \left|\vec{P}_{mc} - \vec{P}_i\right| - \Delta_{xc} - \Delta_{ic}\right) \Big/ \lambda_o$$

Then the equations of condition become:

$$F.3 \quad U_{it} = \lambda_o (N_{it} + N_{ic}) + \Delta_{xt} + \Delta_{it} = \left|\vec{P}_{mt} - \vec{P}_x\right| + \left|\vec{P}_{mt} - \vec{P}_i\right|$$

These simultaneous equations can be solved exactly for three stations or by least squares if there is redundancy of data. The linearized equations are formed by using the linear term of a Taylor expansion about an approximate point. The equations for the approximate position are:

$$F.4 \quad U_{io} = \left|\vec{P}_o - \vec{P}_x\right| + \left|\vec{P}_o - \vec{P}_i\right|$$

Then the linearized equations become

$$F.5 \quad U_{it} = V_{it} + U_{io} + \nabla U_{io} \cdot \left(\vec{P}_{mt} - \vec{P}_o\right)$$

where V_{it} is a correction (termed a residual) to be determined in the range sum. The corrections $(\vec{P}_{mt} - \vec{P}_o)$ to be applied to the approximate position may now be evaluated by methods described in detail in references (2) and (3).

Velocity components may be evaluated either by differencing position data, by differentiating a polynomial fitted to the position data or by solving three or more total derivative equations for $\vec{P}_{mt}$:

$$F.6 \quad \dot{U}_{it} = \nabla u_{it} \cdot \dot{\vec{P}}_{mt}$$

Range difference equations may be formed by differencing the range sum equations. Let $\vec{P}_d$ be the location of a master differencing station.

Then the equations of condition become:

$$F.7 \quad U_{it-d} = \lambda_o (N_{it} + N_{ic} - N_{dt} - N_{dc}) + \Delta_{it} - \Delta_{dt} = \left|\vec{P}_{mt} - \vec{P}_i\right| - \left|\vec{P}_{mt} - \vec{P}_d\right|$$

The approximate point $\vec{P}_o$ may be determined by solving three simultaneous equations exactly. This is most easily done by working in a coordinate system relative to the transmitter. The equations of condition now become:

$$F.8 \quad U_{it} = \left|\vec{P}_{mt}\right| + \left|\vec{P}_{mt} - \vec{P}_i\right|$$

These equations may be linearized by subtracting $\left|\vec{P}_{mt}\right|$ from both sides and squaring:

$$F.9 \quad (U_{it} - \left|\vec{P}_{mt}\right|)^2 = \left|\vec{P}_{mt} - \vec{P}_i\right|^2$$

The resulting equations may be solved for $\vec{P}_{mt}$ by methods detailed in reference (2).

In the case where a master differencing station is used in a range sum-range difference solution, the approximate point is most easily computed in a coordinate system relative to the master station, treating the transmitter as a receiver in the equations.

G. POSSIBLE IMPROVEMENTS IN THE UDOP SYSTEM

Two of the greatest advantages of the UDOP system are its relative simplicity and its low cost. The chief liability appears to be its high dependence upon the human factor during data processing. The early portion of the doppler record must be read manually, and a large number of human decisions must be made concerning reduction time intervals and cycle count corrections. This being true, most of the possible improvements in the system appear to lie in the field of engineering design rather than in data processing. The possible engineering improvements include the following:

(a) the use of phase-locked receiving equipment

(b) digitization at the receiving sites

(c) the use of a VLF reference signal to enable the extension of baselines

(d) increasing the 50KC clock frequency to 200 KC or higher

(e) the use of an offset frequency to enable automatic reductions through nulls

(f) the recording of beat frequency data during a test

(g) the use of a higher reference frequency

If the use of a VLF reference should prove practicable and if the present precision of measurement were maintained, it would be possible to measure the velocity of an earth satellite to an accuracy of better than one

foot per second at any point in the usual orbit. The extended baselines would also eliminate the need for continuous acquisition of data from lift-off and would allow UDOP data to be used for the geodetic positioning of remote stations. The positioning could be done most accurately if data from a wide variety of trajectories were available. The equations to be solved for $P_1(x,y,z)$, using the nomenclature of section F would be:

$$\text{G.1}\quad \lambda_c (N_{it} - N_{ic}) + \Delta_{xt} + \Delta_{it} - \Delta_{xc} - \Delta_{ic} - \left|\vec{P}_{mt} - \vec{P}_x\right| + \left|\vec{P}_{mc} - \vec{P}_x\right| = \left|\vec{P}_{mt} - \vec{P}_1\right| - \left|\vec{P}_{mc} - \vec{P}_1\right|$$

where $(N_{it} - N_{ic})$ is the number of cycles counted from time c until time t. The equations (G.1) can be solved by the same methods as equations (F.3).

As a result of recent investigations, it is now apparent that the UDOP reduction process can be greatly improved by utilizing beat frequency data. This type of data can be conveniently used in a range sum-range difference (one range sum and two or more independent range differences) solution which offers the following advantages over the standard range sum solution:

(1) Random errors in position and velocity data can be reduced significantly.

(2) More accurate estimates of position and velocity random errors can be achieved.

(3) Editing of the digitized raw data can be greatly simplified, allowing more rapid reduction.

(4) Operating time on DARE can be cut by a factor of from two to four.

(5) Anomalies in the digitized raw data can be resolved by inspecting oscillograph playbacks.

(6) Data from noncoherent station arrays can be utilized, adding to system accuracy.

Advantage (1) is true for two reasons. First, the low frequency beat data can be digitized with greater precision on DARE than the high frequency range sum data. Secondly, the geometric dilution of precision for the range sum-range difference solution is normally less than that for the range sum solution.

Advantage (2) is true because correlated errors in the range sum data are largely removed in the differencing process. This will also allow the equations of condition to be weighted by the variances of the measurements.

Editing becomes more simplified because of the smaller number of cycles encountered. Less time is needed on DARE because digitization can be done at full speed.

At present, only the early portion of a UDOP record can be checked manually. The lower beat frequencies can be checked manually throughout a trajectory. Large errors can also be tolerated in the data from the station whose range sum is used since an error in this data will produce position errors only of the same order of magnitude as the measurement errors.

Beat frequency data can be obtained in the following three ways:

(a) by mixing two frequencies in real time during a test

(b) by mixing after a test and recording the beat frequency on magnetic tape

(c) by mixing two frequencies as they are fed into DARE.

REFERENCES

1. The Doppler Automatic Reduction Equipment - Prepared by Electronic Translation Engineering, RCA Service Company, MTP, PAFB, Florida

2. RCA Data Reduction Technical Report No. 42 - A Least Squares Solution for DOVAP Position and Tie-in Point Considering Refraction - James B. Willmann

3. RCA Data Reduction Technical Report No. 39 - A Treatment of Analytical Photogrammetry with Emphasis on Ballistic Camera Applications - D. C. Brown

AN AUTOMATIC DATA REDUCTION FACILITY COMBINING MAXIMUM VERSATILITY AND SPEED

By: Mr. William R. Schumacher, U. S. Navy Underwater Sound Laboratory,
and Mr. H. Malcolm Wilkinson, Epsco, Incorporated

INTRODUCTION

In the field of underwater sound the capability for taking data on a real time basis has always existed. Although the ability to reduce this data also has existed, there has always been a tremendous time lag in the reduction and analysis of the data.

During the past few years there has been a trend toward defining the signal as well as noise in terms of its statistics. To carry out this operation requires a large number of samples of the signal and further increases the time lag and drain on the manpower to reduce the data. In the past, several attempts have been made to automate the reduction of such data. One such attempt consisted of making oscillograms of the analog signal and manually reading the data from the oscillograms. This method was very costly in terms of time and manpower.

An obvious solution to the problem is the construction of a special-purpose digital equipment for processing the data. In general, such equipment is expensive to construct and can handle only a specific set or class of problems. Such a proposal would be economically feasible only if a large volume of data of a particular type existed. Within the Underwater Sound Laboratory, however, the volume of data of a specific

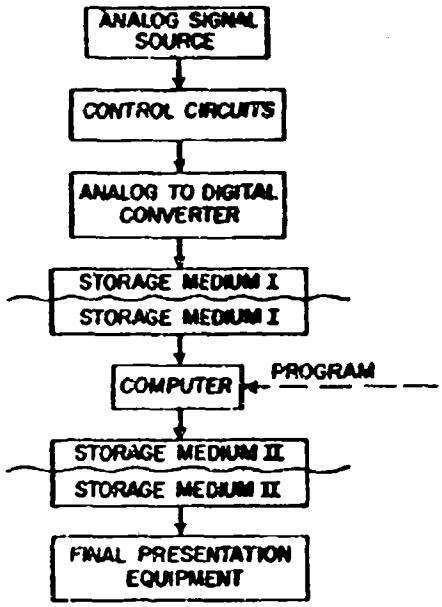


Fig. 1 - Automatic Data Reduction System

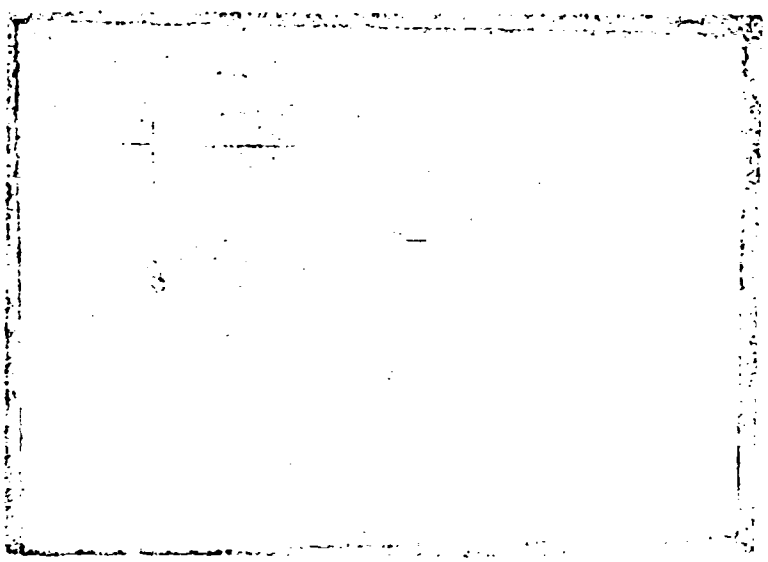


Fig. 2 - USL Analog-to-Digital Conversion System

type as a rule is not sufficiently large to support this solution. Moreover, the various classes of problems that exist at the Laboratory would require many different types of systems. These conditions, then, rule out the consideration of the special-purpose digital equipment from an economical as well as practical point of view.

A possible solution to the problem would be a high-speed, stored-program, digital computer with a means for entering data into the computer for reduction. In the system concept, shown in Fig. 1, the analog information must be converted into digital form and formatted for high-speed entry into the digital computer. The means of entry would be magnetic tape, since it is more compatible than other entry means with the speed of the digital computer. It should be noted at this time that if the analog-to-digital conversion equipment is to handle a large variety of problems, it must approach the generality and speed of the stored-program digital computer. Thus, it must combine high speed with versatility in the number of input analog channels and must have the ability to select a particular portion of input data to convert.

The concept of using a general-purpose digital computer to analyze field data is not new. For quite some time it has been the practice in the business world, as well as in the scientific field, to hand-process raw data and feed this information into a computer for analysis. This system concept utilizes the tremendous speed and versatility of the general-purpose computer with a highly versatile analog-to-digital conversion equipment.

A set of terminal specifications covering such an analog-to-digital conversion system was prepared by the Underwater Sound Laboratory approximately one and a half years ago. The contract for the design and development of such a system was awarded to Epsco, Inc., of Boston. The system as constructed is shown in Fig. 2.

DESCRIPTION OF THE SYSTEM

As conceived the system may be divided into two distinct modes of operation: (1) the Data Acquisition Mode and (2) the Format Conversion Mode. The equipment, however, does not lend itself readily to separation since individual portions may serve a dual purpose; that is, they may operate in each mode. To facilitate the discussion of the system, we shall consider each mode of operation separately.

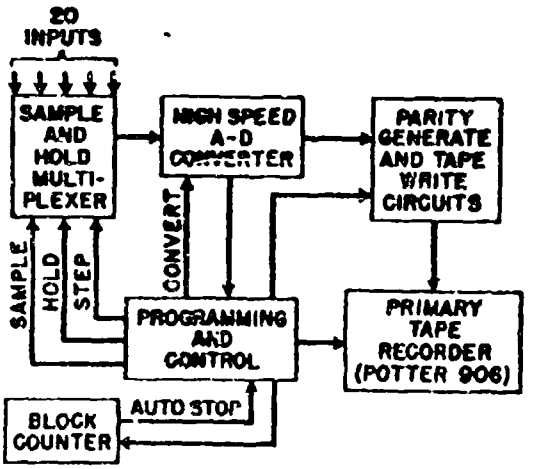


Fig. 1 - Data Acquisition Block Diagram

Data Acquisition Mode

A block diagram of the data acquisition portion of the system is shown in Fig. 3. The 20-input analog voltages are continuously followed by the multiplexer while it is operating in the sample mode. On being triggered by a HOLD pulse, all 20 inputs are disconnected and all input voltages existing at that instant are stored on 20 low-leakage capacitors. Next, the storage capacitors are sequentially switched to the Analog-to-Digital Converter by STEP pulses. Each STEP pulse, after an appropriate delay to allow for decay of switching transients, is followed by a CONVERT pulse transmitted to the A-D converter. After each conversion the digital data are presented to the parity-generate and tape-write circuits and then recorded on the magnetic tape.

As soon as the 20th conversion is completed, the multiplexer receives a SAMPLE pulse. This pulse resets the multiplexer to the sample mode for a small fraction of a millisecond, a sufficient length of time to enable the storage capacitors to acquire charge voltages corresponding to the new input voltages. The multiplexer is then ready for a HOLD command and a series of STEP and CONVERT operations. The timing is arranged so that there is no break in writing the data on magnetic tape. Continuous recording may be used at tape speeds up to 150 inches per second and packing densities up to 240 lines per inch.

The digital data appears in the form of a binary-coded, three-decimal digit number plus sign for each data point. Three lines are re-

corded for each data point. Seven tracks are employed on the 1/2-inch tape. Four tracks are used for numerical data, one track for sign, and one track for parity. The seventh track is used to record a block mark after each block of 20 data points (or 60 lines). This block mark is used later in the tape-to-tape format conversion phase.

Each scan of 20 data points is initiated by the HOLD pulse. The pulse may be triggered internally from a crystal controlled oscillator or from an external scan rate control. This operation can be performed at any speed and need not be at a uniform rate. Flexibility is achieved by the provision for short-cycling the multiplexer. When all 20 channels are used, the maximum number of data points per second for each input is 600. With the short cycling feature the programmer may be set to scan up to 10 channels at 1200 scans per second, 5 channels at 2400, etc., to 12,000 scans, or samples, per second on a single channel. (Note that with a "Sample and Hold" multiplexer no increase in effective sample rate is achieved by paralleling inputs since the same voltage would simply be digitized twice.)

Provision is also made for medium scan rates in which the tape is run at 37-1/2 inches per second. An intermittent or start-stop mode which will handle any scan rate from 60 per second on down is also available. Starting and stopping may be manual or controlled from an external pulse. Stopping may also be controlled automatically by a preset scan counter.

The input multiplexer consists of four Epsco 5-channel EM5IS standard units. These have full-scale inputs of ±10 volts. The transfer accuracy is ±0.05 per cent. The A-D converter is an Epsco Model M617 Datrac. When this unit is used, a complete conversion takes place in 26 microseconds. The primary tape recorder used is a Potter Model 906, which has speed ranges of 37-1/2, 75, and 150 inches per second and a rewind speed of 300 inches per second.

Format Conversion Mode

As previously described, the digital data have been acquired and recorded on the primary tape in blocks of twenty 3-digit decimal numbers. Each block of data occupies 1/4 inch (minimum) with a special block mark recorded at the end of each block, that is, at 1/4-inch intervals. In the Format Conversion Mode the information is transferred to the secondary tape and rearranged in computer format.

The computer installed at the Underwater Sound Laboratory is a Datatron Model 205. The format that is used calls for writing the data in blocks of twenty 10-digit words with sign. A block of data is preceded by a block address. In addition, conditioning codes and parity bits must be generated. This format enables the computer to search for a particular block address before reading out the data. Programming is simplified by requiring that only one data point per word be recorded and that the remaining most significant seven digits be written as zeros.

The preparation of the secondary tape is outlined in Fig. 4. The primary tape is started,

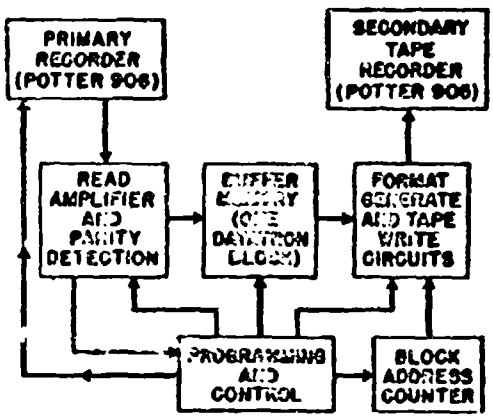


Fig. 4 - Format Conversion Block Diagram

a block's worth of data—20 data points—is transferred to the 240-bit magnetic core buffer memory, and then the tape is stopped. In order to accomplish this operation, the read circuits are enabled at the first block mark, the information is read and transferred to memory, and the read circuits are disabled at the second block mark, which also stops the tape. The signal which stops the primary tape also starts the secondary tape. After a delay sufficient for the tape to come up to speed (5 ms), the block address is recorded from the block counter. The content of the buffer memory then is written onto the tape complete with the required parity bit. As soon as the 20th number is recorded, the secondary tape is stopped.

While recording on the secondary tape is taking place, as already described, the primary tape reverses, senses one block mark, and stops. It is now in a position to read the second block of data into the memory. This operation is initiated by the STOP TAPE 2 signal, which indicates that the first block transfer is complete.

The parity track which was recorded on the primary tape is sensed at the same time that a block of data is being read into the buffer memory. If an error occurs, emptying the buffer onto the secondary tape is inhibited. Instead, the primary tape reverses, senses two block marks, starts forward again, and attempts to read the same block into the buffer a second time. If it succeeds, the routine continues. If it fails, the format conversion operation is stopped and may be started again manually. This operation results in a block of zeros instead of the erroneous data.

The Potter 906, which is used in the Data Acquisition Mode, is also used in the Format Conversion Mode. In fact, the primary tape reel need never be removed from the handler. All that is necessary is that it be rewound and that the data be erased before starting a second acquisition phase. The buffer memory consisted of Epsco magnetic cores. The secondary tape is a Potter Model 905. Epsco standard circuits are used for the read and write amplifiers, digital logic, counters, and control circuits.

SOME PERFORMANCE STATISTICS

Let us now put the performance criteria into terms that are a little more familiar. A standard reel of magnetic tape is 2400 feet, or 28,800 inches. Thus, at 20 data points per 1/4 inch, over 2,300,000 three-digit data values can be recorded on primary tape. If we assume that 300 numbers can be typed on a single page, this is equivalent to over twelve 600-page volumes of data. At 150 inches per second a complete primary tape reel can be recorded in 200 seconds, or 3-1/3 minutes. This is the same as writing a 600-page volume in 20 seconds.

Since a Datatron block of data occupies approximately 3 inches, it takes 12 secondary tapes to store the data recorded on one primary tape. The format conversion process described takes place at approximately 10 transfers per second, or 15 minutes, for a 2400-foot Datatron reel. Preparation of 12 Datatron reels, that is, conversion of the data acquired in 3-1/3 minutes, would take 3 hours. Experience has shown that building in the parity check on the primary tape and the two-pass attempt in the playback mode was well worth while.

Two basic side benefits may be derived from this system. The first is in the construction of auxiliary or portable digital gathering equipment. This portable equipment need not prepare the complicated Datatron 205 tape format but simply the primary tape format. The primary tape may then be formatted to the Datatron 205 via the central processor. The second benefit is that the primary tape, with very little modification to the basic unit, may be entered into an IBM 704 computer.

SYSTEM APPLICATIONS

Let us consider now two different types of problems that may be analyzed by means of this system. The first of these is the analysis of the

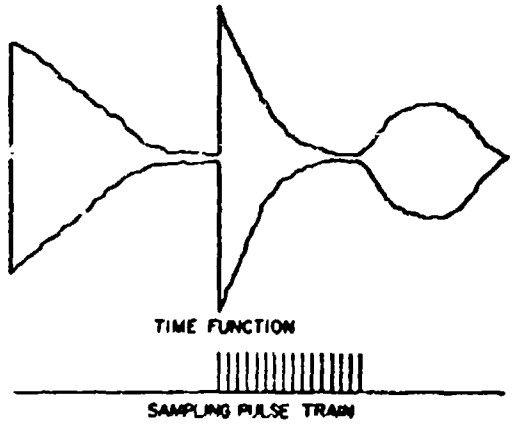


Fig. 5 - Periodic Amplitude Modulated Signal

time function shown in Fig. 5. This time function consists of a carrier, f_c, and a periodic modulating envelope of bandwidth W. What we wish to do is to describe the center portion of this time function in terms of its first- and second-order instantaneous amplitude statistics, as well as the statistics of the envelope and the correlation of the envelope on a periodic basis. In order to accomplish this description without loss of information, we must sample the time function, the Hilbert transform of the time function, and the detected envelope at points 1/W apart for the region of interest.

The procedure requires that the analog information be recorded on one channel of a two-channel magnetic tape unit. On the second channel would be recorded a sampling pulse train, the time orientation of which would be that of the region of interest.

Analog equipment would be used to generate the Hilbert transform and the detected envelope of the signal. This information together with the time function would require three analog channels in the analog-to-digital conversion system.

The system would be operated in the external control mode, with respect to scanning of the multiplexer as well as starting and stopping of the primary tape. The first pulse of the sampling pulse train would start the primary magnetic tape unit. The second pulse and those following it would cause the multiplexer to scan and sample the three inputs. This operation would continue until a predetermined number of scans, corresponding to a certain number of data blocks, have been filled. The scanning operation, as well as the magnetic tape unit, then would be stopped by the preset block counter output. The equipment would then go through the same cycle for the next group of sampling pulses.

In this manner the analog data are sampled, converted to digital form, and recorded on magnetic tape. By external control of the starting and stopping of the magnetic tape unit the data are packed very closely on the tape. Since the time sequence is preserved on the primary tape, the computer may now be programmed to compute the statistics as well as the correlation.

A second application would be the study of a broad-band signal. The aim of such a study is to determine the fluctuations of the spectrum level as a function of time and frequency. The signal would be applied to the input of a comb filter (constant bandwidth). The output of each filter would be envelope-detected and smoothed with two different time constants. The 20 smoothed outputs would be fed to the 20-channel multiplexer.

The equipment would then be operated in two modes, first in a rapid sampling of the smoothed outputs, and a second, at a slow sampling rate. During the rapid sampling mode the tape would run continuously and the sampling or scanning of the multiplexer would be slaved to an internal pulse generator. Because of the "sample and hold" operation of the multiplexer for each scan pulse, samples are made of the 20-input channels.

In the slow sampling mode, the samples are desired once a second. To achieve maximum packing of the information, the primary magnetic tape unit would be operated in the intermittent mode. The pulse to initiate the scan (i.e., 1 per second) would be derived from an external source. This would start the tape unit, and 5 ms later a scan pulse would be initiated. After the conversion of all 20 values, the tape unit would stop. This operation yields a duty cycle of 5 per cent for the tape unit. Both sets of data then would be run through the Format Conversion Phase in order to obtain a magnetic tape for entering into the computer for analysis.

We have attempted in this a[illegible] to obtain an analog-to-digital conversion system of a versatility compatible with a general-purpose digital computer. Since it was not possible at the time of development to specify all problems of data reduction and to include their resultant system demands in the design, the external mode of control was added. Through the external control, additional circuits may be added for each different reduction problem.

THE DIGITRON, A HIGH SPEED DATA DISPLAY SYSTEM

by

Messrs. P. J. Meredith, F. A. Paulus, and D. J. Griffin

The Marquardt Corporation, Pomona Division

INTRODUCTION

The advent of the digital computer and other high speed data handling systems has created a need for compatible readout equipment. The DIGITRON, a high speed, alpha-numeric, visual display device, is designed to fulfill this need. The DIGITRON accepts digitally coded input information from a data processing system and translates this information into characters and symbols which are displayed on a cathode ray tube. The heart of the DIGITRON is the unique character generation and positioning circuitry which converts high speed data into necessary deflection and intensity signals for display on a standard CRT. Significant features of the DIGITRON are:

(a) Solid state circuitry throughout the system
(b) High writing rates
(c) Program control of character size
(d) Readability at high ambient light levels
(e) Wide range of alpha-numeric and special symbols
(f) Easy changeability of symbols
(g) Flicker free display
(h) No special computer program for displaying solid lines

SYSTEM DESCRIPTION

In the application to be described, the DIGITRON is being used as a computer readout device providing a real time, high speed visual display. A parallel 30-bit, 100-microsecond input word is provided: 18 bits are for positioning, 6 for character selection, 3 for line modulation, 2 for size control and 1 for tube selection. Two 19-inch cathode ray tubes are utilized although only one tube is used at a time. The display is switched from one tube to another in one word time. The primary advantage of the DIGITRON, here and in many other applications, is its ability to display data at speeds which are comparable to the operating rates of the input devices. The DIGITRON is required to present characters A through Z, numerals 0 through 9, and several special symbols. The placement of these characters is controlled by 9 bits of X, or horizontal data, and 9 bits of Y, or vertical data. In addition to writing alpha-numeric information, the DIGITRON is required to draw several types of lines—either straight, dashed or tapered. An additional requirement is the ability to circumscribe a character with a circle in the same word time. In a typical operation, the DIGITRON, in one word time, generates a dashed line, writes a character, then draws a circle around that character.

Figure 1 shows the block diagram of the DIGITRON. The main blocks are the stroke timing unit, character selection unit, character storage unit, character generator unit, and position generator unit. The positional data from the computer, moves the electron beam on the cathode ray tube to a particular location. During the beam positioning time, the line modulation as directed by the computer is affected. Once the positioning has been completed, one of 64 possible characters is written. The character writing is done by the stroke timing unit, character selection and storage unit, and character generator unit. Six bits of the input binary word from the computer tells the character selection unit which of the 64 characters to display. Upon receipt of the start or initiate pulse, the stroke timing unit generates sequential wave forms which parallel feed all 64 character cards in the character storage unit. The card selected by the character selection unit produces the necessary digital wave forms to drive the integrators

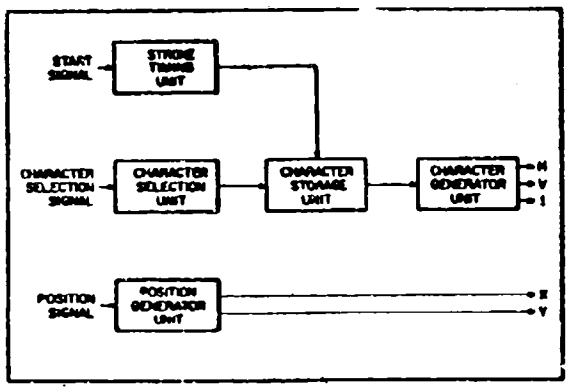


Figure 1. Simplified Block Diagram

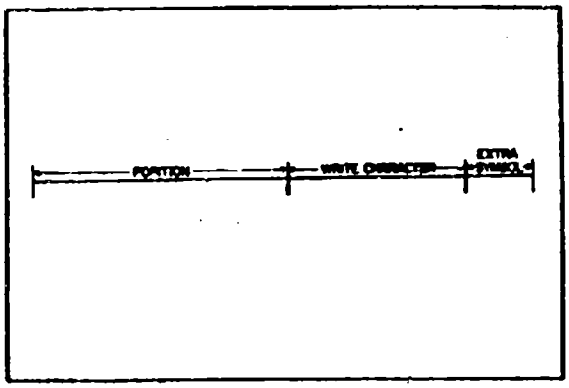


Figure 2. One-Cycle Timing Diagram

and associated circuitry in the character generator unit. The character generator unit then converts the digital inputs to the wave forms necessary to display particular character on the screen for visual presentation.

STROKE TIMING UNIT

One of the more significant problems associated with this application is the speed of operation dictated by the computer word time. This is the time represented by the one cycle timing diagram in Figure 2. In this case the total word time is approximately 100 microseconds: 50 microseconds for positioning, 26 for character writing, and 24 for writing circles. The character writing time of 26 microseconds means the stroke timing unit has to generate a pulse every 1.3 microseconds. These pulses are generated by using the initiate pulse to set a flip flop which starts a free running blocking oscillator. The oscillator generates the stroke timing pulses until the flip flop is reset. Obtaining the necessary response from the associated digital circuits to this repetition rate, requires printed circuitry for the backboard wiring of the stroke timing unit and the character storage unit. This repetition rate also requires that the flip flops in the output buffer of the character storage unit and the intensity amplifiers have response times in the 50 millimicrosecond region.

CHARACTER STORAGE UNIT

The character storage unit contains all the individual character cards. It is the function of each card to select the proper sequence of pulses from the stroke timing unit.

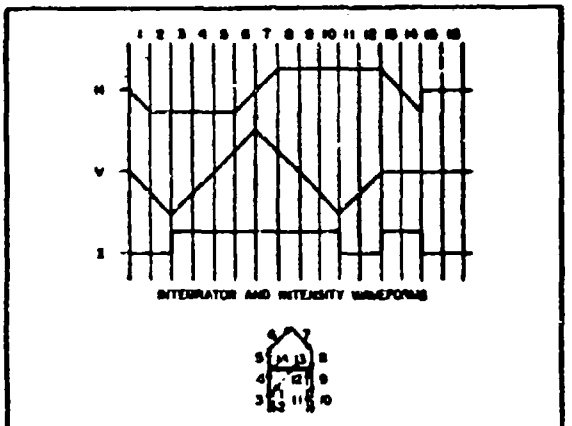
One of the unique features of the character storage unit is that only one character is formed on a removable printed circuit card. These cards can be easily and inexpensively replaced by an identical character card or one that produces an entirely different character. For example, a character card for the letter B can be replaced by a character card which will produce the letter R. This means the entire alphabet can be rearranged or changed to other symbols and special characters, merely by making up these new symbols and characters on universal character cards and then reloading the character storage unit.

It is necessary that all of the flip flops in the output buffer of the character storage unit and those in the stroke timing unit be at a logical zero at the time of the initiate pulse. This is accomplished by forming a pulse on every character card termed the end-of-character pulse. This pulse has several important functions: It stops the generation of the stroke timing pulses, resets all flip flops in the system to zero, and starts a delay circuit which controls the circle generator. The circle generator circuit provides sine waves in a 90 degree phase relation which produces a circle when applied to the plates of the cathode ray tube. In addition to these functions, the end-of-character pulse can initiate another word by signaling the computer that the character writing time is completed. This provides a minimum average character writing time.

CHARACTER GENERATOR UNIT

The function of the character generator is the conversion of the digital signals from the character storage unit into analog deflection voltages. Characters are formed by a series of short strokes. The time base for the strokes originates in the stroke timing unit. Each stroke is of constant time amplitude and ultimately results in proportionately equal deflection increments. This technique is analogous to the familiar Lissajous figures formed with an oscilloscope.

Figure 3 illustrates formation of the character A. The electron beam starts from a central position and is deflected along the path designated by stroke 1. (There is a 1 to 1 time correspondence between the numbered sections of the horizontal and vertical deflection voltages and the stroke numbers on the letter "A" shown in Figure 3.) The beam continues along the path designated



The H & V waveforms combined with the intensity (I) waveform will produce the letter "A" when applied to the electrostatic deflection plates of CRT. The equally spaced vertical lines are timing marks and there is a one to one correspondence between these marks and the stroke numbers shown on the letter "A".

Figure 3. Symbol Generation Timing Chart and Waveforms

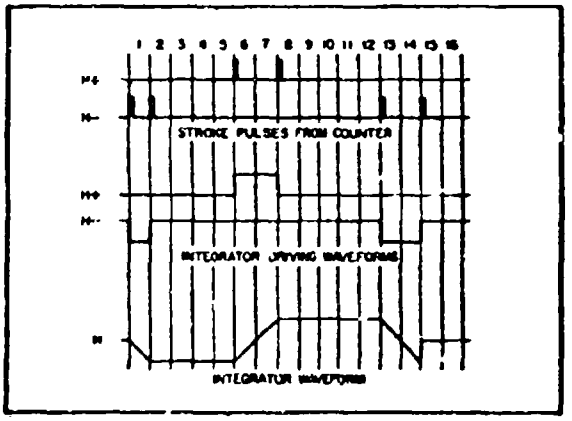


Figure 3A. Symbol Generation Timing Chart and Waveforms

by strokes 2, 3, 4, 5, etc. Note that at time 3 the intensity blanking signal is removed. It is reapplied during times 11 and 12, and again at the end of the character formation time 14.

Figure 3A shows the formation of the horizontal deflection signals. A similar procedure not shown in the figure produces the vertical deflection signals. Digital pulses, +H and -H, are furnished from a selected character card. These pulses ultimately control deflection right and deflection left, respectively. The same pulses

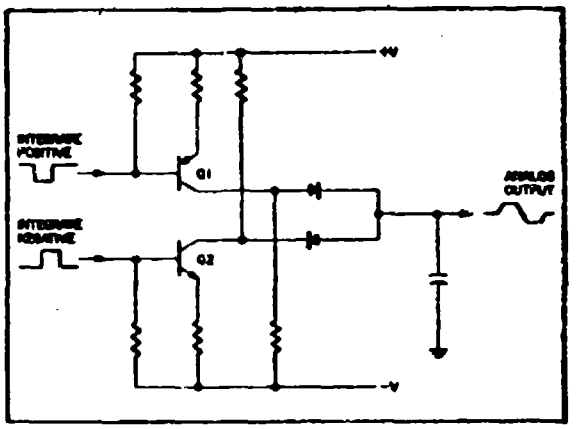


Figure 4. Integrator Circuit

set and reset a flip flop whose outputs control the integrator. The integrator output after amplification is in the necessary form for horizontal deflection.

A simplified integrator circuit is shown in Figure 4. Transistors Q1 and Q2 are connected as positive and negative constant current generators. They serve to charge the output capacitor positively or negatively, depending on which transistor is energized. The diodes, through which the capacitor is charged, are back biased in the absence of charging current. This prevents the capacitor charge from leaking off. Thus the capacitor serves as a memory element as well as a component of a ramp generator. This capacitor is DC restored by a clamp transistor (not shown in the figure) after each character has been written. Character size is digitally controlled by attenuating the integrator output voltages with bilateral transistor switches. The digital size control is followed by a manual size control.

A character size of two inches and the deflection sensitivity of the cathode ray tube require a maximum of 400 volts peak to peak for electrostatic deflection. The character writing speed of 1.3 microseconds per stroke requires about a one megacycle bandwidth. The circuit shown in Figure 5 was designed to meet these requirements. A step-up transformer is used to achieve the final voltage output. Diffused junction silicon transistors with 100-volt breakdown ratings and a cutoff frequency of eight megacycles are used in the amplifier. The amplifier design allows a transformer step-up ratio of only 1 to 2-1/2 to be achieved by providing an 80-volt output at the transformer primary winding. The amplifier also provides the low output impedance necessary to drive two transformers in parallel. This feature eliminates the necessity of having to switch the amplifier output from tube to tube. Two step-up transformers are used to decrease the effect of the stray wiring capacitance associated with the secondary winding. This capacitance reduction is achieved by locating each transformer close to the deflection plates of each tube. In order to provide symmetry, and hence minimize deflection defocusing, the deflecting field should be produced

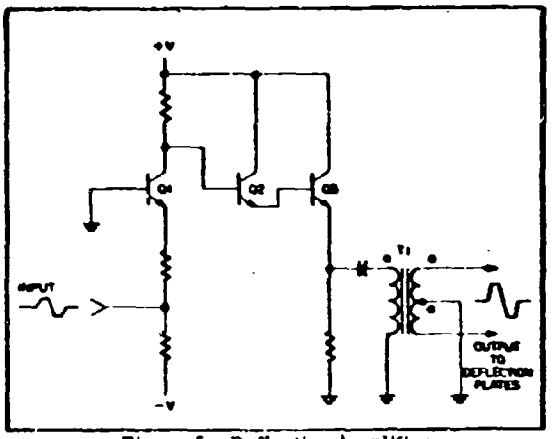


Figure 5. Deflection Amplifier

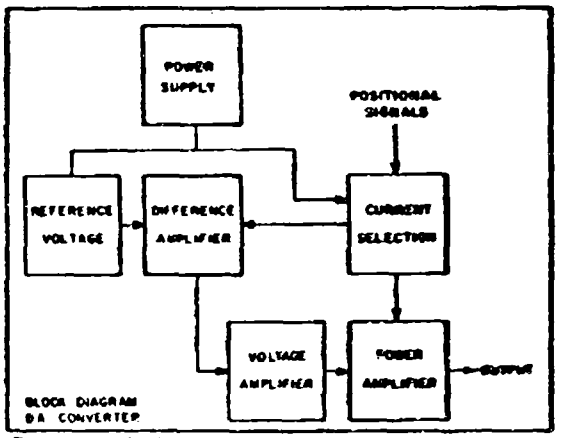


Figure 6. Block Diagram Digital to Analog Converter

so that the mean potential of the parallel plates remains constant. This is accomplished by the center tapped secondary of the step-up transformer. Additional symmetry is gained by balancing the capacitance of each secondary winding to ground and to the primary by a special winding technique. A simplified deflection amplifier schematic appears in Figure 5. A grounded base stage, Q1, with local negative feedback provides voltage gain. An output impedance of 100 ohms is achieved with the cascaded emitter followers, Q2 and Q3. The last follower transistor operates Class A with a power dissipation of approximately 16 watts.

POSITION GENERATOR UNIT

The primary position of the electron beam in the display tube is accomplished magnetically by the position generator unit. This is done by controlling currents in a precision deflection yoke. These currents are proportional to the positional input signals of which there are a total of 18, 9 bits for X deflection and 9 for Y deflection. Since these input signals represent a 9-bit binary word in each axis, it is possible to position the electron beam to any one of over 260,000 positions. In addition to controlling the final position of the electron beam, the deflection yoke also controls beam movement. That is, by forcing any change in yoke current to be linear and requiring that change to occur during a preset time interval, the electron beam will move in a linear manner from point to point. This technique allows the generation of lines between any two points on the face of the tube.

The deflection yoke itself consists of four windings with nearly equal inductance in each winding. Two windings are used for X deflection and two for Y. The yoke windings are phased so that equal currents in all windings cause the electron beam to be in the center of the screen. Whenever the current in one yoke winding is increased the current in the opposing winding is decreased by an equal amount. That is, if we designate the X windings as X_1 and X_2, then if the current in X_1 is increased by one ampere, the current in X_2 would be decreased by one ampere. This mode of yoke operation has several advantages: It allows the use of a single power supply, provides a constant load, and reduces by one-half the current that one driving circuit must handle.

The block diagram in Figure 6 is a simplified version of the complete positioning circuitry required to drive a single yoke winding. This circuit must provide three functions: Convert the digital voltage inputs to precise analog current equivalents, control the rate of change of yoke current, and provide sufficient voltage to drive the yoke winding. The digital to analog conversion is obtained by switching binarily weighted resistors in and out of the current selection circuit shown in Figure 7. There is one such circuit for each bit. The weighted resistor is the resistor shown in the emitter of Q3. The transistors Q1 and Q2 in Figure 7 generate constant currents which charge and discharge the capacitor C_t linearly. This voltage wave form applied to the base of Q3 causes the emitter current to change linearly from a maximum value to a minimum value or vice versa. Constant voltage is maintained across the current selection network by a difference amplifier and the feedback loop. The magnitude of this voltage is determined by the reference voltage which is obtained from a Zener diode circuit that allows a ±10% variation in the primary power supply. The outputs of the current selection circuits are bussed together and connected to the power amplifier. Since a constant voltage is maintained across these circuits, the total current that flows into the power amplifier is proportional to the parallel sum of the binarily weighted resistors.

A tetrode transistor is used in the power amplifier to handle the required 0.05- to 4-ampere current excursions. This 4-ampere current range is equivalent to about a 10-microampere change in the output of the differential amplifier. The emitter of the power tetrode is connected to the current summing bus, and it is to this junction that the feedback loop is connected. The collector of the tetrode then adjusts to whatever voltage is necessary to provide the correct emitter current

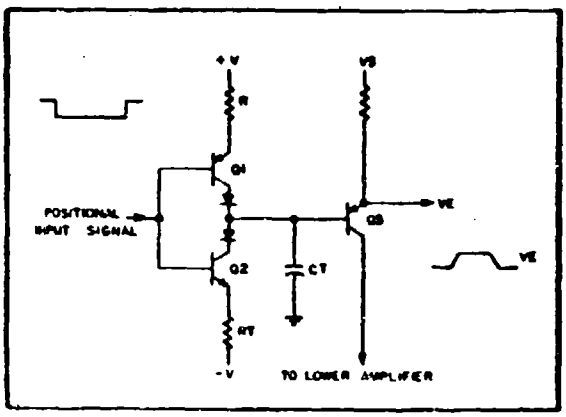


Figure 7. Current Selection Network

as dictated by the feedback loop. The voltage amplifier provides the necessary voltage gain to drive the power amplifier, and a considerable degree of isolation between the differential and power amplifiers.

The operation of the deflection circuitry in Figure 6 is as follows: If the voltage being sampled by the differential amplifier changes, a signal is developed and transmitted to the voltage amplifier, then to the power amplifier in a sense to restore the sampled voltage to its original value. There is normally about 10 volts available to the power amplifier to maintain or change the current in the yoke winding. Though this application of the DIGITRON requires two display tubes, only one driving circuit for each yoke winding is provided. The output of the driving circuit is switched from one yoke to the other on receipt of a signal from the computer. This signal activates a transistor switch which routes the driving circuit output to the correct yoke.

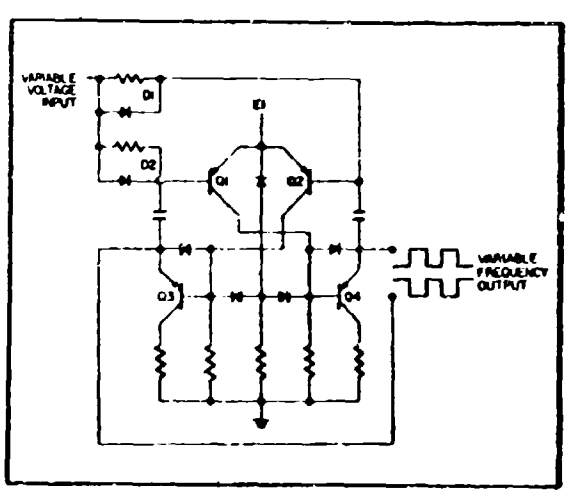


Figure 8. Voltage Controlled Oscillator

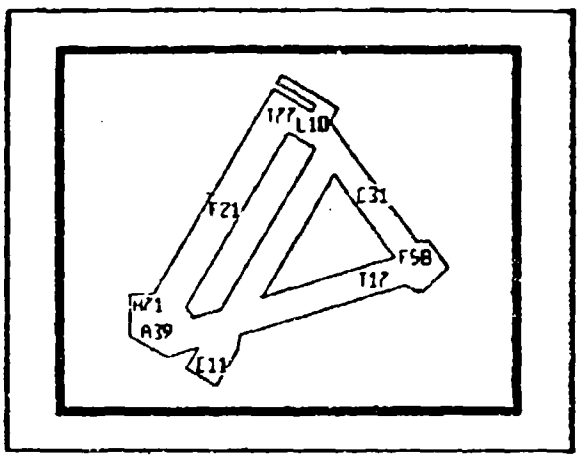


Figure 9. Typical Digitron Display

Associated with the position generator unit is the voltage controlled oscillator circuit (Figure 8.) The function of this circuit is to provide a frequency which can be used to modulate the lines generated by the position generator unit. This modulation frequency is used to control the intensity amplifier during the beam positioning time. The modulation provides the dashed lines as directed by the 3-bit modulation code from the

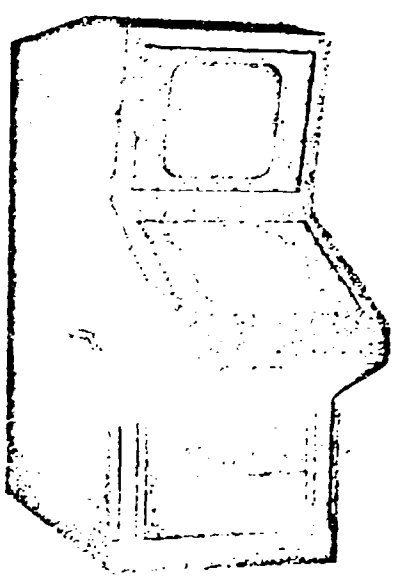


Figure 10. The Digitron

computer. Since the beam positioning time is a constant, but the distance the beam must move is a variable, a constant modulation frequency cannot be used. For example, if a dashed line is required, with dashes equal to 1/4-inch in length with an equal spacing of 1/4-inch between dashes, a frequency that would be a function of the total distance to be moved would have to be generated. This is accomplished by controlling the frequency of an oscillator with a voltage that is proportional to the total distance the beam moves. This voltage is an analog signal that is supplied by the computer. The oscillator itself consists of a free running multivibrator with the timing components connected to the variable input voltage. The wide frequency range is obtained by a combination of transistors with a high beta and a low base emitter drop, and solid state voltage sensitive capacitors.

CONCLUSION

A description has been given of a high speed, solid state data display system which accepts data from a digital computer and provides a wide range of symbology easily read under high ambient light levels.

The development of the solid state circuits in this system has resulted in a significant decrease in the magnetic deflection times previously obtained with similar vacuum tube circuits. This development also demonstrates that the high voltages and high speeds usually associated with electrostatic deflection can be achieved with transistors.

The reliability of the DIGITRON is demonstrated by the recording of over 400 hours of trouble-free operation.

Due to the modular design approach employed, the DIGITRON can be utilized in a wide range of applications. Continuous visual monitoring of process controls, computer program checkout, air traffic control, tactical systems and tabular format presentations are but a few of the applications of the DIGITRON.

A SATELLITE MICROWAVE TELEMETRY OSCILLATOR USING TRAVELING-WAVE TUBE TECHNIQUES

H. R. Johnson
L. A. Roberts

Watkins-Johnson Company
Palo Alto, California

Introduction

Space vehicle telemetry systems are being forced into the microwave range by the need for wider bandwidths and higher effective radiated power. At microwave frequencies, tubes can be built which will meet the bandwidth requirement while highly directional antennas become feasible to meet the radiated power requirement. This order of magnitude upward step in frequency has disclosed a lack of suitable microwave devices. These devices must be capable of generating transmitting power levels while at the same time meeting a number of other requirements peculiar to space vehicle systems. Power levels in the ten-watt range at 2000 mc are beyond the capability of present solid state devices, so the use of a microwave tube is dictated.

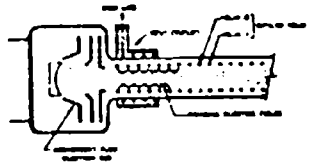
The primary requirements of a space vehicle system are long unattended life, high efficiency, lightweight and extreme reliability. A year of continuous operation is only a beginning; ultimately 20,000 to 40,000 hours and more will be required. High efficiency is of prime importance because of the severe power supply weight penalty per delivered dc watt to the tube. When the total power supply is considered including solar cells, storage batteries and dc converters, the power cost is several pounds per watt. This, of course, subtracts directly from the pay load. The more efficient the tube, the less power supply weight will be needed.

Description of the Tube

To meet these requirements a traveling-wave tube amplifier-oscillator is under development. It incorporates a number of features which will allow it to achieve the following characteristics: 10-watts power output at 2200 mc, life greater than 10,000 hours, efficiency greater than 20 percent, weight less than one pound and rugged construction to prevent microphonism during acceleration, shock and vibration. As an oscillator, its frequency can be voltage-tuned ± 25 mc.

The tube is a helix-type traveling-wave amplifier. The amplifier can be made into a highly stable voltage-tunable oscillator by using external feedback from output to input through a stable microwave cavity resonator. In either case, good efficiency is achieved because the tube is basically an amplifier. In the oscillator case, the voltage tunability requirement puts tight stability requirements on the power supply. For applications where frequency stability is extremely stringent, the over-all satellite system will probably generate a highly stable driving signal and use the tube as an amplifier. For less stringent frequency stability applications, over-all system simplification should accrue through the use of the tube as an oscillator.

To meet the high efficiency-lightweight requirements of this application, a form of non-dissipative electron beam focusing must be used. Two forms of non-dissipative focusing exist: permanent magnet and electrostatic. Of the two, electrostatic focusing involves less weight and is completely insensitive to ambient temperature variations. Thus, electrostatic focusing was chosen. The specific embodiment involves a bifilar helix which can support a voltage difference between the individual helices. A solid beam of circular cross section is shot into the beginning of the helix with a convergent flow electron gun as shown in Fig. 1.



The electrostatic lenses formed by the fringing electric fields between the helix wires produce a net inward focusing force on the electron beam. This force just counteracts the outward force due to mutual repulsion of the electrons. The electron beam can be focused from the cathode to the collector electrode with very low interception of current on any electrode. Beam current transmission from cathode to collector greater than 99 percent has been demonstrated with this tube.

Coupling of energy onto the helix at the input and off the helix at the output is accomplished by the use of coupled helix directional couplers. The coupling helix is a simple filar helix which matches directly to 50 ohm coaxial transmission line. The coupling helix is located outside the vacuum envelope and provides a simple system of coupling to the bifilar helix while simultaneously obtaining voltage isolation from the inner helices.

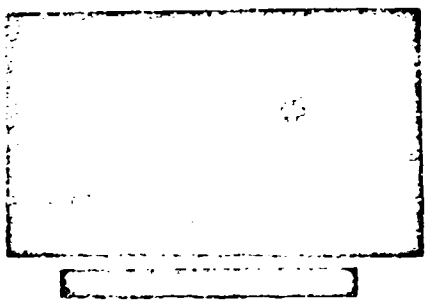
The dissipation of the beam power collected in the collector electrode is by means of a heat sink. The collector electrode will be operated at ground potential so that it can be thermally connected to the heat sink. All other electrodes will then float above or below ground as is necessary to perform their various functions.

The tube is mounted in an aluminum capsule. One side of the capsule has flat surfaces which bolt to the heat sink of the space vehicle. Internally, the collector electrode is thermally bonded to the capsule and the rest of the tube is potted to the capsule walls so that the vacuum envelope of the tube can have little temperature difference with respect to the heat sink.

The weights of the component parts of the encapsulated tube are as follows:

Tube	2.8 oz
Cavity	2.0 oz
Capsule	5.0 oz
Total Weight	
As an amplifier	7.8 oz
As an oscillator	9.8 oz

A photograph of the tube is shown in Fig. 2.



Frequency-Modulated Oscillator Operation

As an oscillator, a fraction of the output power which the tube is delivering to the load is coupled back to the input coupler through a tuned coaxial cavity resonator. The resonator, which is thus located in the feedback loop, determines the frequency of oscillation. The regenerative feedback due to the amplifier in the feedback loop makes the effective Q of the cavity many times greater than its normal value and thus the oscillator can become extremely stable at a fixed value of helix voltage. A cutaway drawing of the oscillator capsule is shown in Fig. 3.

A useful feature of this type of oscillator is its voltage tunability. Within the pass band determined by the loaded Q of the cavity, the frequency of the oscillator can be varied by changing the helix to cathode voltage. For a loaded cavity Q of 40, the oscillator can be tuned ± 25 mc with less than 3 db power output variation. Linearity of the tuning curve is determined by

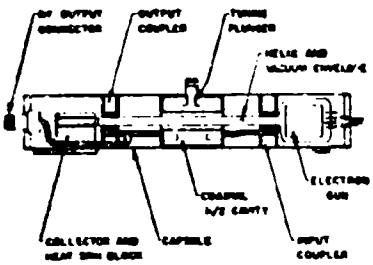


FIG. 3 — CUTAWAY VIEW OF ENCAPSULATED SATELLITE OSCILLATOR SHOWING TUBE, COUPLERS AND CAVITY.

the frequency deviation required. A tuning slope of 0.5 mc/volt is a typical value. Where deviations of only ± 100 kc are required, the cavity Q can be increased by a factor of 10 and the tuning slope will become 0.05 mc/volt. This latter case reduces the stability requirements on the power supply.

The center frequency about which the oscillator can be voltage-tuned is determined by the mechanical tuning of the cavity. The cavity is a half-wave coaxial resonator made of Invar which is further temperature-compensated to make its frequency independent of temperature. The cavity is mechanically tunable across the range of 2000 to 2400 mc by a plunger which perturbs the E field maximum at the cavity center. Coupling into and out of the cavity is accomplished by means of large coupling loops projecting into the cavity from the end walls. By adjusting the coupling, the loaded Q of the cavity is set to the desired value.

Efficiency

Efficiency of 30 percent (including heater power in efficiency calculation) is obtained by (a) designing a tube with a good basic beam efficiency, and (b) by using the efficiency improvement technique of operating the collector electrode at a voltage depressed to approximately one half the average helix voltage.

It should be noted that this 30 percent efficiency is many times greater than other voltage-tuned microwave oscillators because the tube is basically an amplifier and thus can have the efficiency demonstrated by microwave amplifiers. Typical backward-wave oscillators and reflex klystron oscillators have efficiencies in the range of 0.1 percent to 1.0 percent.

A basic beam efficiency of 20 percent is obtained by designing the tube to operate in the high impedance dispersive region of the helix phase velocity characteristic. This region of the helix characteristics can be used because of the narrow band requirements of the system.

Another important aspect of high efficiency operation

to good focusing. A few milliamperes interception on the high voltage helix will cause an appreciable degradation of efficiency performance. Ninety-five percent to 98 percent transmission of cathode current to the collector is necessary.

With care taken in the collector electrode design to suppress secondary electrons emerging from the collector which will dissipate their energy on the helix, the collector can be depressed to less than one half the average beam voltage. A typical distribution of power dissipation on the various electrodes and the resulting efficiency is shown in Table I.

Efficiency

Heater power	3.5 watts
Collector power	6.6 watts
Low helix	.01 watt
Anode and "Body Collector"	33.2 watts
High helix	5.2 watts
Total input power	48.6 watts
Output power	15.0 watts
Efficiency	30.9 percent

TABLE I

Estimated power dissipation on tube elements and efficiency

Life

Operating life of a traveling-wave tube is primarily a function of the cathode life. As long as the cathode emits electrons in a uniform manner across its face and is operating under space-charge limited conditions, the tube will continue to deliver its rated power output. Actually it will continue to perform well beyond this point. However, when cathode emission begins to change under standard conditions, it is a signal that end of life is imminent.

Several factors lead to long cathode life all of which are incorporated in this tube. The cathode is designed to operate at 136 ma/cm^2 current density. At this current density, cathode temperature can be kept in the range below 700°C. With passive nickel used in the cathode and the low temperatures of operation cited, reaction rates of barium oxide breakdown to pure barium in the cathode can be kept to low values. This insures an adequate supply of free barium at the cathode to maintain emission for a period in excess of 10,000 hours.

This supply of free barium would not be adequate in the presence of even a small amount of ion bombardment of the cathode. To achieve this long life, ion bombardment must be eliminated. The electrostatic focusing system accomplishes this. Ions generated in the helix and collector regions of the tube do not meet the focusing conditions for stable beam flow. Instead they are drawn radially outward to the least positive of the two helices of the bifilar helix and are swept out of the beam. The ion bombardment of the cathode is virtually eliminated. The remaining source of ions is due to interception of beam current on gun electrodes and the first turns of the helix, where voltage gradients in the gun would draw these ions into the cathode. Good attention to beam focusing through this region eliminates this source.

Reliability

The trouble areas which most greatly affect tube reliability are voltage breakdown within the tube and heater failure. Through careful design, the maximum voltage gradient across any insulator within the tube is 13 kv/inch. This is a conservative figure. All insulators across which voltage gradients exist are protected from evaporated metal deposits which can increase chances of surface breakdown.

Heater failure possibilities are minimized by sintering the heater into the cathode structure. This makes the heater-cathode a compact thermally-joined unit and reduces temperature difference between heater and cathode by approximately 300°C. The heater operates much cooler thereby reducing the embrittlement of the wire, the tendency of the heater coating to crack during thermal stress, and the chance of heater breakage due to vibration.

Other factors which bear on reliability such as choice of materials, processing techniques and care in construction of the tube are rigidly controlled.

Design for Environmental Conditions

This tube is designed to survive the shock, vibration, and acceleration conditions that will be encountered when the equipment in the satellite is passing through the launching phase. The further rigid requirement is imposed that it must also be capable of operating while in this launching environment without microphonic degradation of performance.

Microphonic modulation of a traveling-wave tube can result from primarily three things: First, motion of the helix which can change its propagation properties with time; Second, relative motion of the electron beam and helix which will change beam helix coupling with time; and Third, relative motion of electron gun electrodes so that beam current will change its magnitude or position with time. Each of these possibilities have been eliminated in the basic construction philosophy of the tube.

The helix is supported by a structure of ceramic rods where each turn of the helix is glazed to the rods. The helix is rigidly locked down and cannot move. Its propagation characteristics cannot change with time. The cage of rods is in turn supported by a tightly fitting stiff glass envelope.

The beam is focused by the helix itself and thus cannot have any relative motion with respect to it. This would not be true if focusing were provided by an external magnetic focusing structure. If tube and magnet had any relative motion at all, microphonic modulation would result.

The electron gun is built in a very rugged fashion with gun elements made of thick metal pieces. These parts are held in place by a cage of rods which run parallel to the axis of the tube and are glazed to them. The gun is supported on one end by mating into the helix structure which guarantees permanent alignment and by the base of the tube on the other end. The parts are thus prevented from having relative motion. The cathode is supported by a stiff conical metal sleeve.

The entire tube is rigidly supported in its capsule so that relative vibration of tube and capsule parts cannot occur.

Conclusion

A microwave tube has been described which will meet the requirements of the output stage of a satellite born telemetry system. It is capable of an operating life greater than 10,000 hours. It has an efficiency capability of greater than 30 percent including the heater power. The total weight of the encapsulated tube is less than one pound. Table II is a list of the significant operating parameters and characteristics of the tube.

Frequency range	
Mechanical tuning	2000-2400 mc
Electronic tuning	± 25 mc
Power output	15 watts(min.)
Voltage	
Collector	Ground potential
Cathode	-315 volts**
Helix 1	850 volts*
Helix 2	2650 volts*
Anode and "Body Collector"	1750 volts*
Cathode current	52 ma
Tuning rate	0.5 mc/volt
Efficiency(including heater power)	>30 percent
Heater voltage	6.3 volts
Heater current	.56 amp
Weight	7.8 oz(as an amplifier) 9.8 oz(as an oscillator
Focusing means	Electrostatic

** with respect to ground
* with respect to cathode

TABLE II

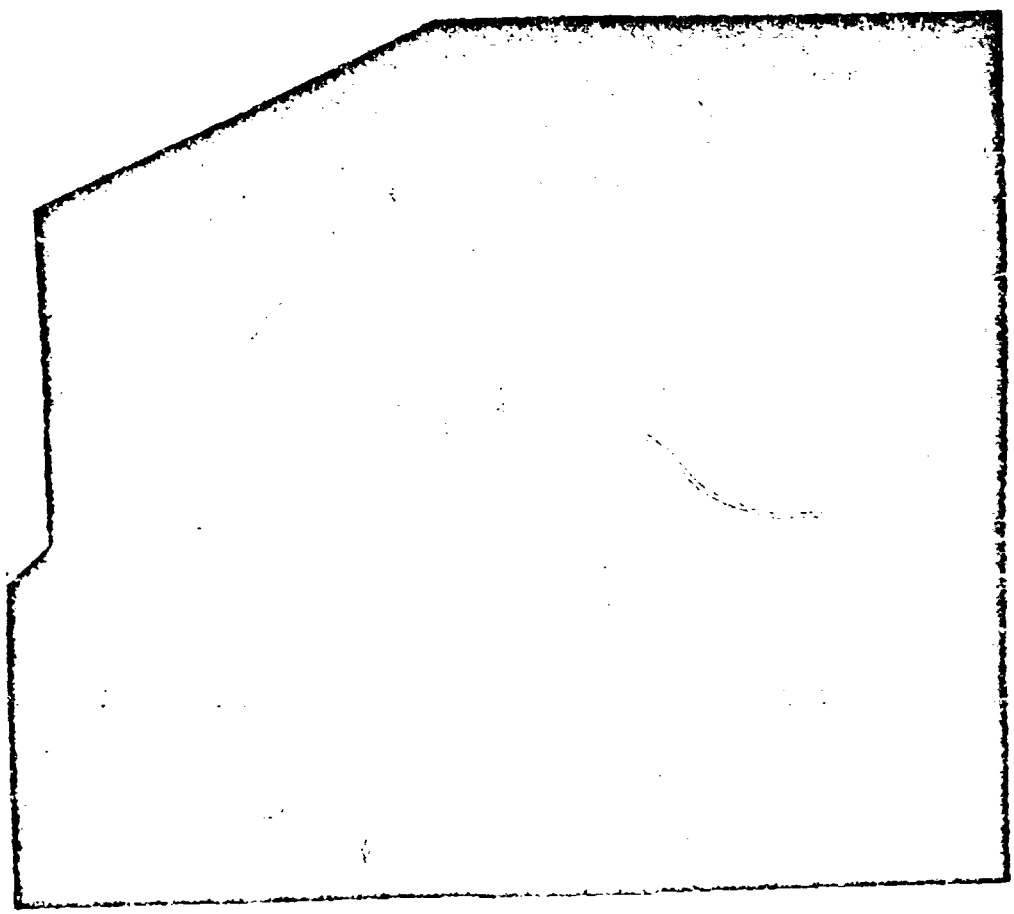
Operating parameters and characteristics of satellite telemetry amplifier - oscillator tube

Acknowledgment

This work is supported by the U.S. Air Force, Air Materiel Command, under WADD Contract AF 33(600)-38302.

SATELLITE RELIABILITY ACHIEVED THROUGH COMPREHENSIVE ENVIRONMENTAL AND FUNCTIONAL TESTING

John A. Chambers, ABMA, Huntsville, Ala.



INTRODUCTION

On 13 October 1959 the Army Ballistic Missile Agency launched its fourth and most complex artificial earth satellite, 92.3 pound Explorer VII. It was orbited by a four-stage Juno II missile with the three upper stages spin-stabilized at 450 rpm. The project was sponsored by the National Aeronautics and Space Administration.

The results of the launch are well known, and will not be elaborated on at this time. It will be recalled that the spent fourth stage rocket was separated as planned, the 20 megacycle turnstile antenna system unreeled at a controlled rate, and all seven of the scientific experiments on board functioned normally.

While extremely gratifying, the success of the satellite was not a surprise to those familiar with the exhaustive test program which accompanied all phases of its development. Even the flight models themselves went through a modified environmental test series corresponding with launch and orbit conditions.

COMPONENT TESTS

Structural. The double truncated cone design was selected as the optimum configuration for strength, ease of manufacture, adequate solar cell area, and convenience of achieving a disk-like mass distribution. The latter characteristic was necessary so that spin stabilization of the body in orbit might be obtained.

A metal shell for the satellite was undesirable because of eddy currents which would be induced by motion in the earth's magnetic field. It was anticipated that with a metal shell, spin and consequently spin stabilization would be lost before the useful lifetime of the satellite expired.

Laminated Fiberglas was selected as the shell material. Shells were built and tested while other phases of the design progressed. A static load was applied along the thrust axis, and stresses were measured until buckling failure occurred. The buckling pattern closely followed the results of a mathematical analysis which preceded the design. It is of interest to note that after removal of the load the shell snapped back into shape, apparently as good as new.

A structural prototype was then tested under conditions more closely simulating the actual launching, by use of a rocket sled. Realistic spin of the payload was accomplished throughout the acceleration, and again test results were in good agreement with theory.

Separation Device. During the above work other features of the satellite were being developed and tested. An important component was the separation device, which freed the satellite from the spent last stage. The separation was essential because the satellite with rocket shell attached did not have a disk-like mass distribution and would have begun to tumble after a relatively short time. It was necessary to provide a small relative separation velocity to remove the shell from the vicinity of the satellite, yet it was required that the separation impulse should impart no disturbance to the spinning body. The separation device was required to be a highly rigid joint under heavy bending loads, which could occur as a result of transverse vibrations of the vehicle and also from possible nutation of the upper stages.

Three different separation device designs were tested in vibration, bending, and actual operation under conditions which gave a good approximation of the dynamics of the actual separation in space.

A dummy satellite with correct mass distribution was built. It was designed so that it might be suspended upright from a cord attached to a ball bearing at its center of gravity. The body was thus free to move in any lateral direction, move up, and to tilt 14° about any perpendicular to the spin axis. The satellite half of a separation device was fitted to the lower end of the dummy, and the rocket half was mounted on a simulated fourth stage shell.

With the separation device joining the two bodies, the upper end of the suspension cord was attached to a crane hook and the assembly was lowered to a spin stool. The stool was designed to impart spin through friction with the lower end of the simulated rocket.

When a spin rate of 450 rpm was achieved, the assembly was lifted a few inches and a separation signal was transmitted through a circuit attached to the suspension cord.

In order to prevent damage to the dummy shell or the spin stool, the latter was mounted on a cart which was propelled away just before separation, drawing taut a safety net behind it.

The motion of the cart, which was on rails, was also utilized to induce nutation in the assembly. An adjustable arm fixed to the cart could be made to strike the dummy shell with the desired interference. Thus nutations were achieved which exceeded by a safe factor the maximum expected nutation after injection into orbit.

The separation device which was finally selected for use was tested 15 times under the above conditions without any malfunction. High speed camera observation of each event showed no disturbance of the simulated satellite at separation.

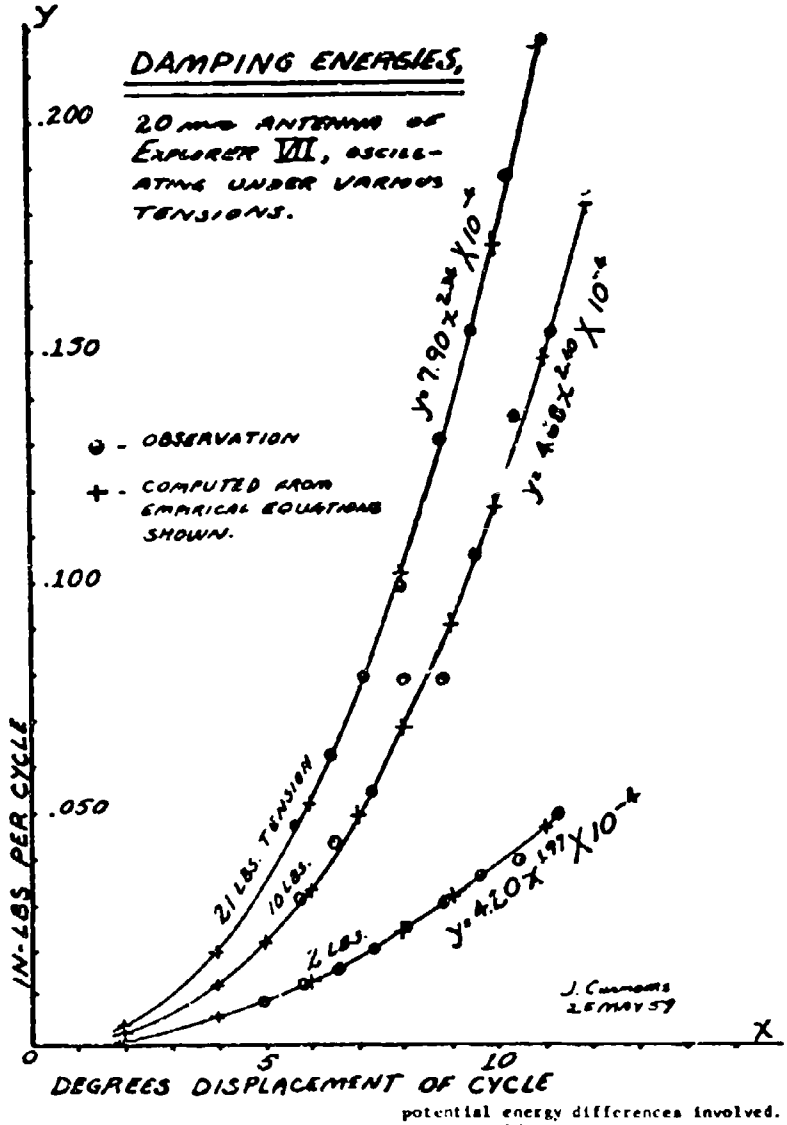
Briefly, the device which was selected for use may be described as a six-segment C-ring clamp. The segments were drawn against the upper and lower mating lips by radial arms engaged in a central gas cylinder assembly. Actuation of the device was by two electrically fired squibs in redundant array. Separation of the bodies was achieved by the expansion impulse of a short spring.

20 mc Antenna. Another factor which was extremely important to the success of the mission was the 20 mc antenna system. A turnstile antenna with each arm equal to 1/4 wavelength (about 12 feet) was specified for the 20 mc telemetry transmitter. The transmitting surface was required to be #16 AWG or equivalent. The most feasible design within reasonable weight limits was a system of flexible wires to be stored within the satellite during the ascending phase of the trajectory. After injection the wires would be unreeled from the spinning satellite at a controlled rate, and would be maintained in radial extension by centrifugal force.

In addition to its function in the telemetry link, the antenna system was required to have physical characteristics which would result in relatively high rates of energy dissipation when flexed. Relative motion between satellite and antennas, as might result from satellite nutation at separation from the fourth stage, would thus dissipate energy until the body returned to true spin.

It was decided that the antennas should extend from the satellite through single-step telescoping Teflon tubes. The telescoping feature increased the initial centrifugal force on the antennas, thereby assuring adequate breakaway force, but did not increase the final centrifugal force as larger weights on the antenna tips would have done. The tubes served the further purposes of distributing the flexing of the wire over a large radius at its point of emergence from the satellite shell, and of providing a larger energy dissipation rate under flexing than the bare wire would have done.

Energy dissipation rates of several types of wire were measured under tensions up to and including the expected centrifugal force on each

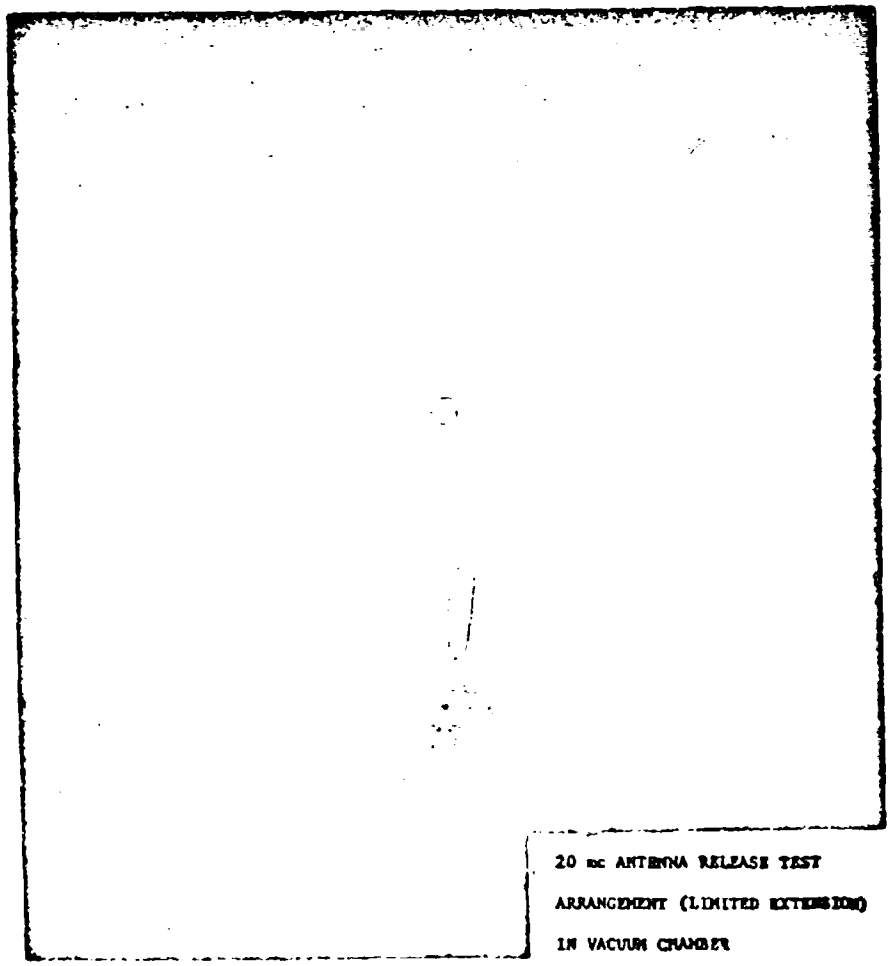


wire in orbit.

A simple pendulum arrangement mounted in a vacuum chamber was utilized for the energy measurements. A six-foot pendulum was assembled with the cable under test comprising the tension member. The upper suspension point included a Teflon stiffening tube which enclosed the cable in orbital configuration.

For each test the pendulum was actuated at an angle corresponding to the maximum expected nutation angle. The amplitude decay rate was correlated with the displacement angle, and a consideration of the system geometry plus the known mass of the pendulum weight showed the potential energy differences involved. Thus curves could be drawn representing energy dissipation loss per cycle for the full range of displacements observed. Integration of the curve which reflected the proper cable tension for orbital conditions gave the total energy dissipation for a given oscillation decay. Comparison with the known excess energy content of a certain nutation condition, as well as the nutation frequency, gave the time required for the antenna system to return the satellite to true spin.

Bending fatigue tests were performed to insure that the cable would not break before



20 mc ANTENNA RELEASE TEST

ARRANGEMENT (LIMITED EXTENSION)

IN VACUUM CHAMBER

damping any oscillation which might occur. Two cable and tube assemblies withstood 3,600,000 deflections of 30° before failure, which occurred in each case at a point of attachment to the test machine rather than at the point of maximum bending.

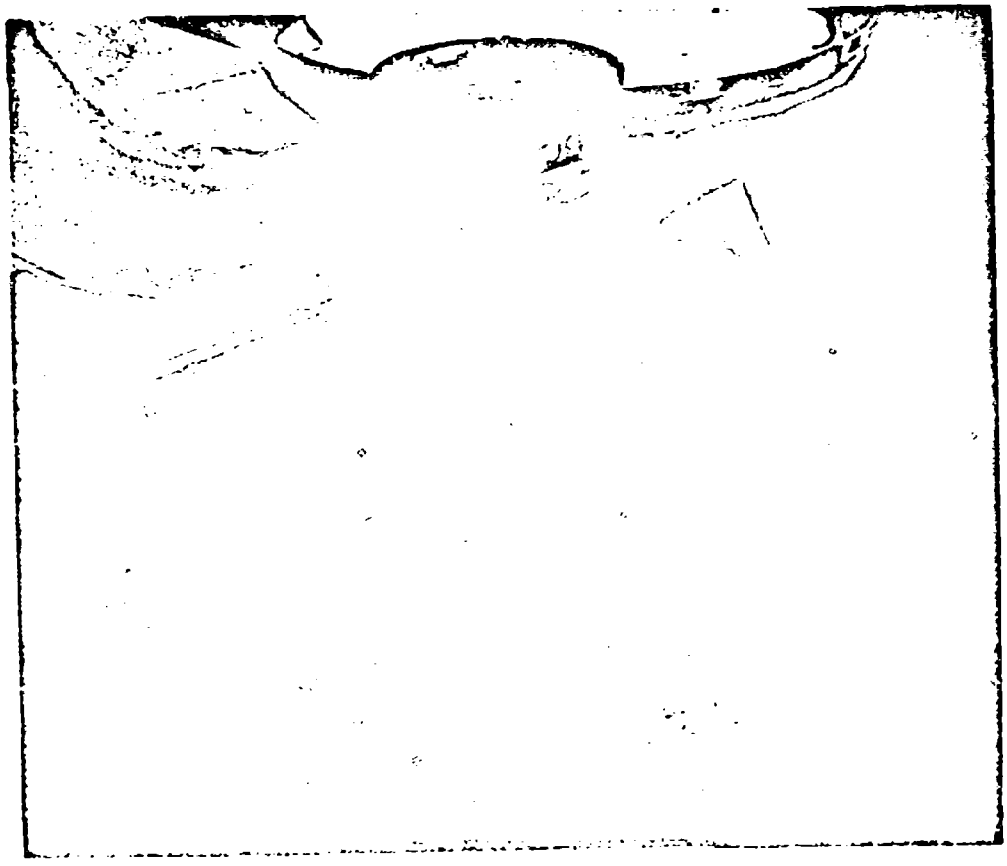
Tension tests over a period of weeks precluded the possibility of excessive elongation or a creepage failure in orbit.

The cable which was accepted comprised a twisted Fiberglas core with a braided conductive layer covered in turn with an extruded Teflon jacket. The cable was 0.085 in. outside diameter

The antenna cable having been selected, testing of the release system proper could begin. This system included an integral ring gear and spool, with separate winding grooves for each arm of the antenna; a reduction gear train; and an electric motor. The ring gear-spool fitted over the central instrument column of the satellite and was mounted in the equatorial plane. A small negative torque was exerted by the drive motor until the release signal was received, in order to prevent creeping out of the antennas under the combined effects of centrifugal force and vibration. During release the amount of wire remaining on the spool was monitored and at the proper time the motor was stopped and the transmitter switched to the antenna system from a dummy load.

Tests of the release system began with releases from a fixture rigidly mounted to a spin stool, in air. While short wires had to be used because of atmospheric drag, the use of compensating masses on the antenna tips permitted proof that the mechanism operated properly under centrifugal force.

Next, a mass-correct fixture was constructed similar to the one used for fourth stage separation tests. Again the fixture was designed for



suspension at its center of gravity, but without a bearing. In this case spin was imparted to the body, through the cord, from an overhead motor and magnetic clutch. As before, the fixture possessed five degrees of freedom, and could tilt 14° about any axis.

Incorporated in the overhead assembly were sockets where weights might be held magnetically, to be dropped when desired. The weight, striking the spinning fixture, induced nutation. In this way observations of the dynamic behavior of body and wires could be made under the worst expected conditions of release in space.

For safety, tests were first performed in vacuum with limited release of the wires, with and without nutation. These tests being successful, seven full-scale tests were run in vacuum, with excellent results. The nutation-damping ability of the system was found to be approximately as predicted, allowing for the self-damping characteristics of the simulated satellite which were observed without antenna release.

PROTOTYPE TESTING

At this time the first prototype was completed. Environmental testing of individual electronic packages and power supply components had accompanied their development. Radiation studies had been made, using mockups similar to the actual satellite. Now tests of the complete satellite in essentially final form could proceed.

Thermal Investigations. Extensive thermal experiments were performed on a prototype to insure that a passive system of temperature control could be designed to maintain the temperature of sensitive components within prescribed limits.

Eleven experiments were performed with the prototype in vacuum of better than 5×10^{-6} mm Hg in order to determine internal conductive and radiative exchange characteristics. For these tests the prototype was chilled to a low temperature, usually -20°C, and a heat step was

applied either to the entire satellite or to selected areas by close-fitting electric heater blankets. The primary inner surfaces of the shell or parts of the shell were raised some 80 or 90 Centigrade degrees and held at a constant temperature by varying the blanket voltages.

Absorption and emission characteristics of the external surfaces were determined by simple laboratory tests, and this information was used in conjunction with the evaluation of the internal characteristics. The passive thermal controls resulting from the tests were: polished gold foil on the central column, polished aluminum foil inside the equatorial band, polished aluminum foil on the battery packs, titanium dioxide paint on the inside of the cones, sandblasted aluminum foil on the exposed Fiberglas of the exterior, and Kokide on the solar cell frames The foil was applied in segments to minimize eddy currents and the space between external foil segments was covered with gray paint. A redesign of battery box supports was required by the results of the thermal investigations. Temperature control of Explorer VII in orbit has been excellent.

Environmental Testing. An environmental prototype was completed and a test program started which, insofar as practical, simulated launch and orbit environments to which the satellite would be subjected.

All new antenna release system components having been installed, it was considered desirable to check gear friction by monitoring drive motor current under variable force such as would be experienced in orbit. Accordingly, a device was constructed which employed a weighted chain drawn over a sprocket by gravity. This apparatus permitted the antennas to be pulled out by a drawbar arrangement, closely simulating the increasing centrifugal force of actual release under spin. The satellite was, of course, held stationary for this test.

After careful dynamic balancing, the prototype was spun at 600 rpm and instrumentation was monitored.

The instrument column was subjected to thirty 25 g shocks without damage.

The prototype was subjected to random noise vibration in the thrust direction and on two perpendicular transverse axes at 20 g, 13 g, and 13 g, respectively. Instrumentation was monitored. As a result of the vibration tests the rivets fastening the lower cone to the separation device were increased in size and the battery box mounting was redesigned. The latter change was thermally unsatisfactory (see above) and was once more redesigned and vibrated. An investigation was made of vibration amplification at various points on the prototype, with particular attention to the solar cell assembly.

An acceleration of 40 g was applied to the prototype for three minutes on the centrifuge. Due to significant gyroscopic forces which would have resulted from spin of the satellite about its own axis, it was permitted to remain fixed. Later the instrument column alone was spun with centrifuge rotation.

The prototype was placed in a vacuum chamber and instrumentation operation was checked at stabilized instrument package temperatures of 0°C, 30°C, and 60°C. Pressure during the checks was about 6×10^{-2} mm Hg.

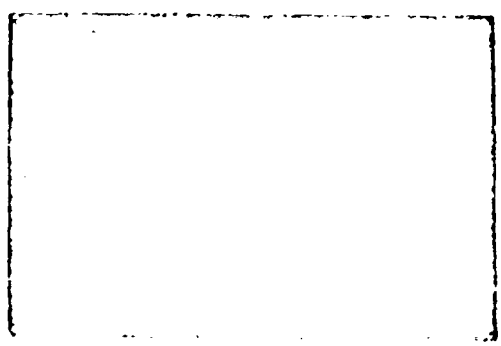
Extensive RF checks were performed to conclude the environmental test program. Field strength measurements were made, and it was determined that transmission from beneath the shroud was adequate for tracking during the ascent of the missile before shroud separation.

Flight Acceptance Testing. Both flight models were subjected to test programs which were essentially the same as the environmental prototype test program described above. Some reductions were made in the severity of mechanical tests to avoid fatigue prior to launching.

CONCLUSIONS

It is apparent from the successful operation of Explorer satellites orbited to date that there is considerable merit in a philosophy of virtual 100% environmental testing of all satellite components, including flight hardware and final configurations. While this approach will necessarily be modified as satellites become larger and components become more standardized, the experience being obtained now will be a major factor in the development of future design and testing parameters.

oOo



A COMPACT UHF DIPLEXER FOR APPLICATIONS INVOLVING ROCKETS OR SATELLITES

By: Mr. Sam E. Parker, Hughes Aircraft Company

I. Introduction

This compact UHF diplexing unit was developed to permit a 100-watt transmitter and a sensitive, low-noise receiver to operate simultaneously with a single antenna while carried by a high-altitude rocket. In this manner, radio-relay service can be provide over greatly extended distances.

The major objectives in the design of this diplexer are:

a. Adequate operational isolation is required between the transmitter and receiver.

b. Insertion loss between the transmitter and antenna should be kept as low as practicable, to conserve valuable transmitter power.

c. Attenuation in excess of 20 db at the transmitting frequency is required for receiver isolation.

d. Space and weight must be held to a practical minimum, consistent with reliable performance aboard the rocket.

These performance objectives were required with the transmitting and receiving frequencies separated roughly ten percent. (For initial studies, frequencies of 240.8 Mc and 264.7 Mc were assigned.)

II. Operational Isolation

The prevention of excessive interaction during tuning operations of the associated transmitter and receiver is a major requirement in a system of this type. It would be quite impracticable to attempt to specify the different values of impedance presented by a class C power amplifier and a communications receiver during typical adjustment procedures. This is particularly true of a class C amplifier whose output impedance is both nonlinear and nonanalytic.

In recognition of this situation, a practical solution to the equipment interaction problem is afforded by a consideration of limiting conditions which include, of course, an open- and a short- circuit. Rephrasing the problem with reference to Fig. 1, the load impedance of the transmitter operating at frequency f_t should remain reasonably constant as the 2-2 port is alternately open-circuited and short-circuited. Similarly, the impedance of the equivalent source supplying a signal to the receiver at frequency f_r should not change appreciably as the 2-2 port

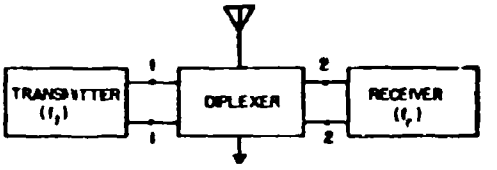


Fig. 1. Generalized diplexing system

is opened and shorted. Denoting the difference between the open-circuit and short-circuit impedances of a linear, passive network by ΔZ, we may write

$$\Delta Z = \frac{2 Z_I}{\sinh 2a} \qquad (1)$$

where Z_I is the image impedance and a is the attenuation constant. If a is 3 nepers or larger (roughly 26 db or more), ΔZ is generally quite small, as indicated by the approximation

$$\Delta Z \approx \frac{4 Z_I}{e^{2a}} \qquad (a \geq 3). \qquad (2)$$

This approach, of course, requires some knowledge of Z_I at the frequencies of interest in the attenuation bands of the diplexing network.

III. Combining Network

In rocket and satellite applications, it is apparent that valuable transmitter power should not be wasted unnecessarily. It follows, therefore, that the transmitting branch of a diplexing unit should not be designed to provide excessive attenuation at the receiving frequency or else excessive losses at the transmitting frequency are bound to result.

In consequence of the foregoing, a special combining network was formulated, as illustrated in Fig. 2. The technique is quite comparable to that employed by Otto J. Zobel in his fractional-π terminated filters;[1] but our problem is somewhat simplified by the fact that we are concerned primarily with only two operating frequencies.

[1] U.S. Patents No. 1,557,229 and 1,557,230, both issued 13 October 1925.

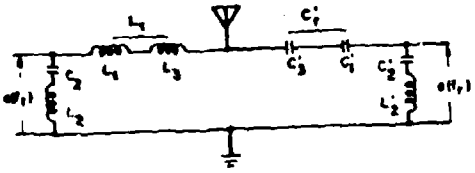


Fig. 2. Basic combining network, assuming lossless elements

Referring to Fig. 2, we assume that the antenna provides a constant, resistive load R_L at frequencies f_t and f_r. In this particular application we may further assume that the network should present this same value of resistance to both the transmitter and the receiver at their respective operating frequencies. Thus, in this case, the equivalent transmitting and receiving networks will each be symmetrical ($L_1 = L_3$ and $C_1' = C_3'$) and the following relationships will apply:

$$\omega_t^2 L_2' C_2' = 1 \tag{3a}$$

$$\omega_r^2 L_2 C_2 = 1 \tag{3b}$$

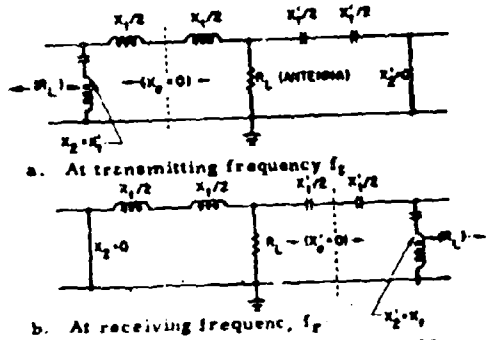
$$-\frac{1}{\omega_t C_2'} = \omega_t L_2 - \frac{1}{\omega_t C_2} \tag{3c}$$

$$\omega_r L_1 = \omega_r L_2' - \frac{1}{\omega_r C_2'} \tag{3d}$$

$$\omega_t L_1 = \frac{2R_L^2}{\omega_t C_t' \left(R_L^2 + \frac{1}{\omega_t^2 C_t^2}\right)} \tag{3e}$$

$$\frac{1}{\omega_r C_t'} = \frac{2R_L^2 \omega_r L_t}{R_L^2 + \omega_r^2 L_t^2} \tag{3f}$$

Perhaps Fig. 3 may be helpful in visualizing these circuit conditions. The expressions (3a) and (3b) denote resonance in the alternate shunt arms, of course, and (3c) and (3d) insure equality of the shunt arms of the symmetrical π-networks. The final pair of equations, (3e) and (3f), denote the fact that the impedance seen at each half of these π-networks must be purely resistive (X_e or $X_e' = 0$), when it is assumed that the transmitter and the receiver present an impedance R_L at their operating frequencies. Although this last assumption is not realized in practice, as previously explained, it leads to the precise transmission characteristics desired.



a. At transmitting frequency f_t

b. At receiving frequenc, f_r

Fig. 3. Combining network simplified for different operating frequencies

From the foregoing, it is clear that the network of Fig. 3(a) is capable of very low-loss transmission at f_t and of considerable attenuation at f_r. The situation is just the reverse in Fig. 3(b). While beyond the scope of this discussion, methods for extending the analysis to include reactive loads or to effect nominal impedance transformations are self-evident.

IV. Bridged-T Network

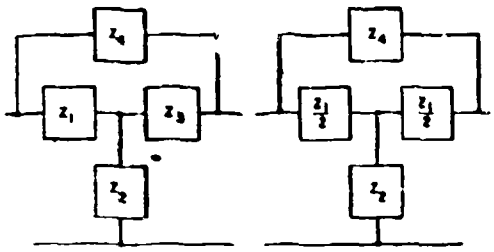
Although the special combining network can provide ample operational isolation and low insertion loss so desirable in the transmitting arm of the diplexer, additional attenuation is required at the transmitting frequency in the receiving arm. In problems of this type, experience has shown that the bridged-T network can provide important advantages.[2] Accordingly, efforts were directed toward the development of a model suitable for UHF applications.

Being a true null circuit, bridged-T networks are capable of infinite attenuation when dissipation inherent in all realizable network elements of a passive nature is taken into account. The expressions which establish a null or balance condition may, of course, be derived in a variety of ways. Probably the most direct approach is to equate to zero the sum of the currents through the T-network and the bridging arm, as explained by Tuttle.[3] Alternatively, one may equate the open-circuit and the short-circuit impedances of the network. In any event, using the general network configuration in Fig. 4(a), the balance equation is

$$Z_1 + Z_3 + \frac{Z_1 Z_3}{Z_2} + Z_4 = 0. \tag{4}$$

[2] See U.S. Patent No. 2,774,069 for transmitter multiplexing system, issued to the author on 11 December 1956.

[3] W. N. Tuttle, "Bridged-T Parallel-T Null Circuits for Measurements at Radio Frequencies," Proc. I.R.E., Vol 28, pp. 23-29, January 1940.



a. General form b. Symmetrical form

Fig. 4. Bridged-T networks

In the symmetrical case of Fig. 4(b), with which we are concerned in this application, the expression for a null condition becomes

$$Z_1\left(1 + \frac{Z_1}{4Z_2}\right) + Z_4 = 0. \qquad (5)$$

If equation (5) is written in complex form and the real and imaginary parts are equated to zero, the conditions for zero transmission are found to be (since $R_1^2 \ll X_1^2$):

$$R_1 + R_4 + \frac{R_1X_1X_2}{2|Z_2|^2} - \frac{R_2X_1^2}{4|Z_2|^2} = 0 \qquad (6a)$$

$$X_1 + X_4 + \frac{R_1X_1R_2}{2|Z_2|^2} + \frac{X_2X_1^2}{4|Z_2|^2} = 0 \qquad (6b)$$

Turning now to the transmission characteristics of the symmetrical bridged-T, the expression for the image impedance is helpful:

$$Z_I = \sqrt{Z_{oc}Z_{sc}} = \sqrt{\frac{Z_4}{Z_1 + Z_4}}\sqrt{(Z_1Z_2)\left(1 + \frac{Z_1}{4Z_2}\right)} \qquad (7)$$

This is recognized as the image impedance of the symmetrical T-network multiplied by the factor $M = \sqrt{1/(1+k)}$ where $k = Z_1/Z_4$.

If we consider the fact that the operating frequencies are relatively close together, the foregoing expressions suggest the use of the network configuration shown in Fig. 5. At the "pass" frequency, considerable simplification results by making Z_1 small compared both with Z_4 and with $4Z_2$ so that equation (7) becomes approximately,

$$Z_I \doteq \sqrt{Z_1Z_2} = R_L \qquad (8)$$

provided, of course, that Z_1 and Z_2 are inverse reactive arms. Obviously, some of these relationships are inconsistent with the null conditions since equation (5) shows that if Z_1 is small compared with Z_2, then Z_1 and Z_4 will be of somewhat comparable magnitudes at the "stop"

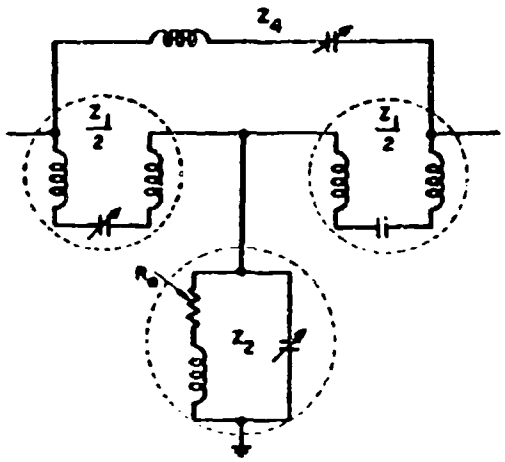


Fig. 5. Bridged-T network form indicated by the desired transmission and rejection characteristics

frequency. A convenient method for alleviating this problem is to design the series arms ($Z_1/2$ in Fig. 5) for large reactance changes with small loss components.

In the UHF frequency range, these considerations practically preclude the use of lumped elements and, in consequence, resonant line sections were employed. The major design considerations will be explained with reference to Fig. 6. For a dissipationless open-circuited line of Fig. 6(a), the input reactance is

$$X_1 = -Z_o \cot \beta s \qquad (9)$$

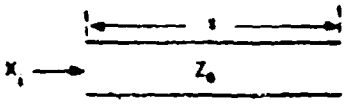
and the slope of the reactance curve is

$$\frac{dX_1}{d\omega} = \frac{Z_o s}{v_p}\csc^2 \beta s \qquad (10)$$

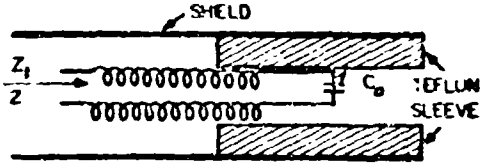
where Z_o is the characteristic impedance and $\beta s = 2\pi s/\lambda = \omega s/v_p$. With the use of odd multiples of quarter-wave sections, the slope of the reactance characteristic is

$$\left.\frac{dX_1}{d\omega}\right|_{\omega=\omega_r} = \frac{Z_o s_r}{v_p}. \qquad (11)$$

For a large slope, therefore, the characteristic impedance and the resonant length of line should be large. Since these considerations are somewhat in conflict with requirements for low dissipation, however, the actual units were constructed with effective lengths near $3\lambda/4$ and with an effective value of Z_o somewhat in excess of 100 ohms.



a. Basic line section



b. Helical resonator

Fig. 6. Development of bifilar-wound resonators employed in series arms

Use of a resonator as indicated in Fig. 6(b) for the series arms may introduce serious problems as a result of the shunt impedance to the shield (i.e., to ground). It is clear that if the open-circuited, two-wire, balanced line provides a low input impedance, it may also provide a low impedance in the coaxial mode, assuming that the electrical lengths are comparable. Use of the dielectric sleeve shown in Fig. 6(b) affords an effective means for solution of this problem, however. After the desired value of Z_1/Z is obtained by adjustment of the added (tuning) capacitor C_a, it has been found to be a relatively simple matter to insert an adjustable Teflon sleeve and approximate anti-resonance (i.e., a half wavelength) in the coaxial mode. Through careful design, relatively little dielectric sleeve is required and the detuning effect on the balanced line is almost negligible.

With regard to the shunt arm Z_2, the design details were influenced by transmission and suppression considerations alike. For transmission at f_t, Z_2 should be largely reactive and the inverse of Z_1, in accordance with equation (8); and for effective suppression at f_s, it is convenient to provide appropriate series components of Z_2 by adding a resistor R_4 as indicated in Fig. 5. In the UHF frequency range, this method for approximating the balance conditions appears to be preferable to the use of series resistance in the bridging arm without adding loss components in any arms of the T-section—an approach which has been found quite satisfactory at low radio frequencies.

Perhaps a numerical example will best illustrate the design procedure which has been followed. At the "transmission" frequency of the bridged-T network (264.7 Mc), each helical resonator was initially adjusted to give an input reactance of j5 ohms (arbitrarily selected), the Teflon sleeves were adjusted for anti-resonance in the coaxial mode, and the resonators were readjusted slightly, as required. The measured value of Z_1 was found to be about 2 + j10. An impedance of 50 - j250 was considered reasonable for the shunt arm Z_2 and was provided by a resistor-inductor combination (about 90 + j330) in parallel with a capacitor (tuned to give -j150). Under the approximations associated with equation (8), the image impedance was roughly 51 ohms. The impedances Z_1 and Z_2 were then determined at the rejection frequency (240.8 Mc) and found to be roughly 6 - j200 and 93 - j304 respectively. When these values were substituted in equations (6), the components of the bridging impedance Z_4 required for a null were computed to be 1.4 + j230.6. These values, of course, were readily approximated by a series combination of a coil and variable capacitor.

V. Physical and Electrical Characteristics

Two views of a complete diplexing unit are shown in the photographs of Fig. 7. This unit has successfully withstood accelerations up to 150 g as well as other environmental tests necessary for high-altitude rocket applications. The entire assembly weighs slightly less than

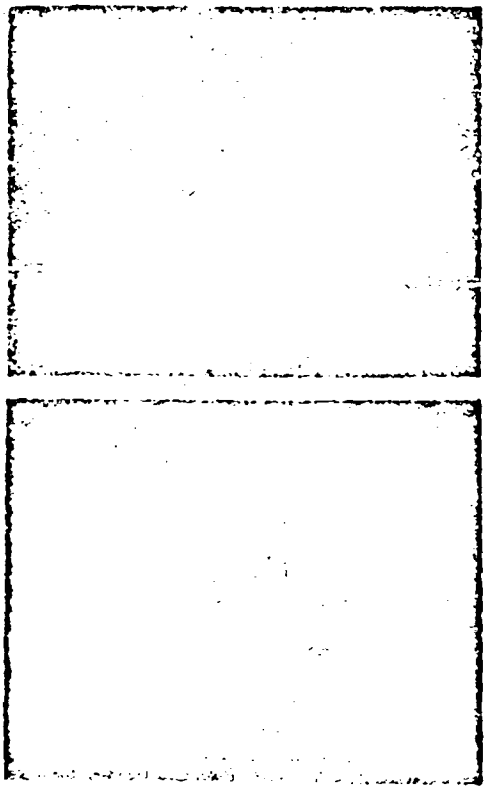


Fig. 7. Photographs of the UHF diplexer

one pound. A complete schematic diagram is included in Fig. 8 with an indication of the compartmentation employed. In addition to the special combining and bridged-T networks discussed previously, the receiving circuit is seen to include an additional π-section patterned after portions of the combining network.

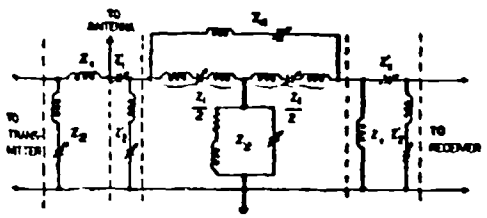


Fig. 8. Complete schematic diagram showing shielded compartments

Measured insertion loss characteristics are indicated in Fig. 9. The insert in this figure gives data on insertion loss and VSWR which are pertinent to this particular application. The value of the attenuation peak at the transmitting frequency was too large to measure with readily available facilities and instrumentation. However, it is known to exceed 100 db.

VI. Acknowledgment

In the experimental phases of this work, the many contributions of Mr. J. F. O'Connor are gratefully acknowledged.

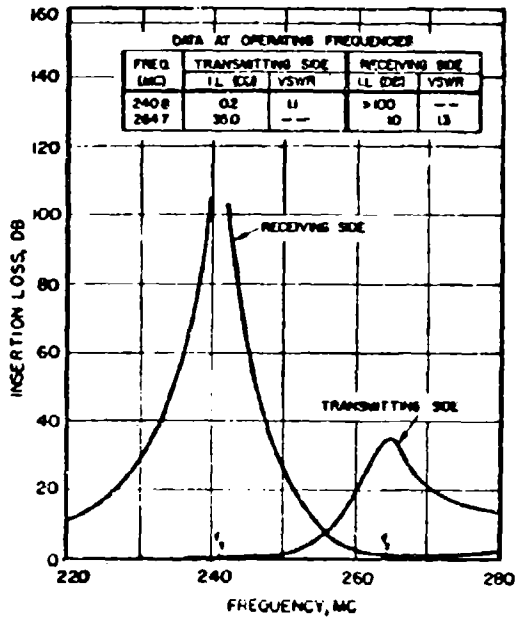


Fig. 9. Measured performance characteristics

STATUS OF ULTRA-LOW-NOISE TRAVELING-WAVE TUBES AND BEAM-TYPE PARAMETRIC AMPLIFIERS

Dr. B. P. Israelsen, Dr. K. L. Kotzebue and Dr. C. E. St. John
Watkins-Johnson Company, Palo Alto

Introduction

In the field of low-noise microwave amplifiers the foremost competitors are (1) the maser, (2) the solid-state diode parametric amplifier, (3) the beam-type parametric amplifier, and (4) the traveling-wave amplifier. The intent of this paper is to discuss the youngest of these devices, the promising and somewhat mysterious beam-type parametric amplifier and the oldest of the group, the venerable but tenacious traveling-wave tube.

This paper begins with a background development of the low-noise traveling-wave amplifier from its inception to recent results on broadband tubes. The possibilities of future improvement and frequency extension are discussed.

The logical development of the beam-type parametric amplifier is presented next with reference to the original work of Adler and Wade. Present design trends are outlined with emphasis on operation over broader frequency bandwidth. A final section discusses some of the systems problems involved in the use of these basically competitive devices. Special attention is paid to the question of effective system noise figure.

Low-Noise Traveling-Wave Amplifiers

In the following an historical background of low-noise traveling-wave tube development will be presented as a basis for evaluating present results and also to allow extrapolation. Implicit in such a presentation is the danger of credit omission. The stress has been on what are felt to be high points in theory or development with emphasis on both individual and company or university contributions.

Early in the history of traveling-wave tube studies L. M. Field and D. A. Watkins, at Stanford University, in a joint invention described a form of electron gun which in varying forms has been a basis for all subsequent low-noise traveling-wave developments. The original form of gun is shown schematically in Fig. 1. By adjustment of electrode potentials within this gun noise performance of the associated amplifier could be varied over wide limits and noise figures as low as 11 db were observed.

In subsequent experimental work C. C. Cutler and C. F. Quate at Bell Telephone Laboratories verified the predicted presence of a standing-wave of noise at microwave frequencies on a magnetically-focused electron beam. The original "velocity-jump" gun of Field and Watkins obviously functioned by control of location and standing-wave ratio of noise on the beam in the region of signal entry to the amplifier. In this manner signal-to-noise ratio and, therefore, noise figure were affected by changes in gun potentials.

Subsequent work by R. W. Peter at the Radio Corporation of America evolved a modified three-region gun, shown in Fig. 1, wherein abrupt potential changes or "velocity-jumps" were eliminated and noise performance of 8 db observed. At this point J. R. Pierce of Bell Telephone Laboratories advanced a theory suggesting a minimum attainable noise figure of 6 db would exist assuming specific values of uncorrelated velocity and current noise at the cathode. In work at Bell Telephone Laboratories one of the authors succeeded in obtaining noise figures of less than 5 db and both theoretical and experimental studies were renewed.

At this point M. R. Currie of Hughes Aircraft Co. in a backward-wave amplifier study employing a modified annular beam measured noise figure below 4 db with a beam potential profile in the gun quite different from those used in earlier studies. In theoretical studies by D. A. Watkins, A. W. Shaw, A. E. Siegman, H. C. Tuieh of Stanford and J. R. Pierce of Bell Telephone Laboratories possible explanations for the reduced noise content of the stream were advanced, based on the concept of noise correlation resulting in a low-velocity drifting beam similar in some respects to Currie's. Predicted noise content for several beams of this character is shown in Fig. 2 with no irreducible minimum value indicated.

While specific reference to their work has not been made, the significant contributions of W. R. Beam, S. Bloom, W. E. Danielson, H. A. Haus and F. N. H. Robinson are hereby acknowledged.

This brings us to a description of recent results obtained at Watkins-Johnson Company on octave bandwidth tubes in S-band. In Fig. 3 the terminal noise figure of a specific tube is plotted as a function of frequency with all voltages fixed. It may be seen that the noise figure is no higher than 5 db at any point with a considerable frequency range for which it remains less than 4 db. A further point of interest is the achievement of a 2.9 db terminal noise figure at one spot in the band.

With a view to assessing further improvements in performance, Fig. 4 shows schematically, several forms of electron guns which emphasize control of beam potential in the critical region of low-velocity near the cathode. The first of these is the type used

originally by Currie. The second gun, proposed by Watkins, uses an array of close-spaced anodes with an annular cathode,to allow even more precise control of beam potential. A gun of this type was used in obtaining the octave bandwidth results just described. A third gun, conceived by one of the authors, uses a small insulated probe centered in an annular cathode to offer further beam control. Obvious combinations of the features shown in Fig. 4 may offer promise as well.

It seems reasonable to conclude that the work to date has not exhausted the possible forms of gun design, offering even lower noise results.

It is interesting to pause at this point and contemplate the possibilities for future improvements and extensions of present low-noise traveling-wave tube results.

In Fig. 6 a series of noise figure achievements have been plotted as a function of time. The plateau-like quality of the plot is interesting as well as the general trend downwards with time. With present theory indicating no limit to noise reduction and the ever present prospect of new techniques arising, it seems safe to predict the achievement of noise figures in the 1 - 2 db region in time.

A further point of interest is to consider the possible extension of low-noise techniques to much higher frequencies i.e., the millimeter wavelength range. In Fig. 7 the lowest noise figures known to the authors at present are shown as a function of frequency. It is apparent that very little variation has been observed in minimum attained noise figures over a 10:1 ratio in frequency from 1,000 to 10,000 mcs.

Extrapolation by a further 10:1 ratio up to 100,000 mcs or 3 mm wavelength, while dangerous, is warranted by virtue of the interest in this frequency range as well as intermediate values. It is believed that the obvious extrapolation from Fig. 6 to approximately 5 - 6 db noise figures at 100,000 mcs would be reasonable if a number of fundamental parameters could be held essentially constant. In particular, the ratios electronic gain/ wavelength, circuit loss/ wavelength, beam plasma frequency/ operating frequency and focusing field strength/ operating frequency are fundamental in this evaluation. In view of the fact that it appears unlikely that all of these ratios can indeed be exactly maintained, an increase in the minimum attainable noise figure is predicted. A best estimate suggests noise figures in the 10 - 12 db range at 100,000 mcs. Correspondingly lower values might be anticipated for frequencies intermediate to the 10,000 and 100,000 mcs limits chosen.

Beam-Type Parametric Amplifiers

In the earlier development stages of the traveling-wave tube many attempts were made to remove beam noise prior to circuit interaction by ingenious coupling techniques. Realization that amplification was achieved in this case by interaction with a "slow"-wave on the beam which is characterized by a negative kinetic energy led to the conclusion that further efforts in this direction would be futile. Nevertheless, a by-product of this thought suggested that noise in positive energy fast waves could be removed from a beam, although needed gain could not be obtained directly by interaction with this wave. At this point R. Adler and G. Wade of Zenith Radio and Stanford University, respectively, evolved the technique of parametric amplification using a transverse-field fast-wave interaction, and in this manner achieved low noise performance.

The basic operation of such a parametric device may be seen best by a block diagram of the form shown in Fig. 5. The input coupler serves the dual purpose of transferring the signal to the beam while reciprocally removing beam noise content. In the next section of the device the energy required for amplification is supplied from an rf energy source or pump. Following amplification, the signal is removed from the beam by a coupler identical in form to that used in the input. Tubes of this type have been reported with double-sideband noise figures of approximately 1 db in the UHF band.

The Adler-Wade device operates with the signal frequency closely equal to the electron cyclotron frequency which in turn is established by the strength of the focusing magnetic field. A result of this mode of operation is that the effective phase-velocity of the desired fast-wave approaches infinity and the use of lumped circuit coupling elements is permissible. The use of lumped elements i.e., a flat plate transverse-field condenser plus a tuning inductance, is delightful from the simplicity aspect but places limitations on the operational bandwidth of these couplers, and therefore of the device as a whole. Fortunately, the pumping structure is in concept almost frequency independent, and this has encouraged studies of broadband transverse-field counterparts of the Adler-Wade device.

A trend in development has been to consider the use of broadband couplers such as various helical structures operating in a transverse-field Kompfner-dip condition. This mode of operation offers identical signal and noise transfer capabilities to the lumped circuit device but with far superior instantaneous and electronically tunable features. Recent studies of couplers of this form at Watkins-Johnson Company to date have yielded signal transfers of approximately 20 db. No noise performance is yet available.

An interesting design feature is the possibility of idler cancellation within the device as a result of using a hollow-beam. While not improving noise performance such a technique would allow operation over extended frequency range without the beating mode of output power observed in the filamentary beam device.

Systems Aspects

In considering possible applications for these low-noise amplifiers it is pertinent to give some general information regarding life and possible environmental conditions. While little specific background on these topics exists for the beam-type parametric amplifier, the close correlation in fundamental design features between the parametric amplifier and the low-noise traveling-wave tube permits the assumption of performance similar to that suggested here on the basis of traveling-wave tube experience.

Briefly, these tubes should have extremely long life capabilities, in the limit approaching those of the submarine-cable type of i-f amplifier. Life ranging from a minimum of 10,000 hours upwards to 50,000--100,000 should therefore be attainable. This opinion is based on two principal considerations. Typical cathode current densities in these devices, particularly at frequencies like S-band, are extremely low and permit operation at temperatures in the 600--650°C range. Since the basic chemical reaction rate for reduction of the oxide cathode materials doubles for approximately each 25°C increase in temperature, a drastic increase in attainable life is to be expected at these temperatures. Furthermore, the tubes can be made almost entirely from refractory materials such as ceramics and molybdenum, allowing stringent heat treatment so that gas evolution during life is minimized. The low beam powers normally used here result in negligible heating of parts during operation so that gas build-up again is limited.

On the subject of environmental conditions the principal problem is that of ruggedizing the rf interaction structure which is typically a helical winding of small diameter wire. By means of longitudinal ceramic support rods to which the helix is intimately bonded, a self-supporting structure of considerable strength and low dielectric-loading can be achieved. This technique, developed first at Bell Telephone Laboratories, has been used successfully in tubes demonstrated as suitable for even missile environments. Upper limits on ambient temperature are commonly fixed not by the tube but by the limits for the magnetic material of the focusing structure where permanent magnets are employed.

It is claimed, therefore, that low-noise amplifiers of the types discussed here can be made completely adequate for operation in modern military environments.

The decision whether to use a traveling-wave tube or a parametric amplifier depends upon the particular system application. The traveling-wave tube is physically a simpler device and requires no high-frequency pump for its operation. Also the traveling-wave tube has broader bandwidth than the beam-type parametric amplifier. While it can be expected that the bandwidth of the parametric devices will be improved, it appears that the traveling-wave tube will always be capable of greater bandwidth. This greater bandwidth is not necessarily an advantage. In applications where jamming must be considered, it is preferable to have a rather narrow bandpass characteristic which can be electronically tuned. In this respect, a parametric amplifier would be preferred. (Other alternatives might be to use a backward-wave amplifier or an electronically-tuned pre-selector filter in front of a traveling-wave tube.)

The outstanding attribute of the parametric amplifier which makes it worthy of consideration is its low noise figure. Omitting masers from consideration, when the ultimate in noise figure is necessary over small or moderate bandwidths, the parametric amplifier can well be the logical choice. We must be very careful, however, and make certain that the parametric amplifier does indeed have the better noise performance. For example, suppose we have a choice between a degenerate parametric amplifier with a quoted noise figure of 2 db, and a traveling-wave tube with a noise figure of 3.5 db. Which will give the best noise performance? Contrary to what one might initially presume, it is the traveling-wave tube that will almost always give the best noise performance. This is because the quoted noise figure of the parametric amplifier is the so-called double-sideband noise figure, which is convenient for the measurement of noise figure in the laboratory, but which cannot be used for a direct comparison of system sensitivity. It is not difficult to make this comparison of system sensitivity if the concept of operating noise figure or operating noise temperature is used. Operating noise temperature will be defined as

$$T_{op} = \frac{N_{out}}{g\,k\,B} \quad \text{where} \qquad (1)$$

N_{out} = total noise power delivered to load

g = transducer gain

k = Boltzmann's constant

B = noise bandwidth

A simple rearrangement will show that

$$S_{in} = (S/N)_{out}\, K\, B\, T_{op} \quad \text{where} \qquad (2)$$

S_{in} = input signal power

$(S/N)_{out}$ = output signal-to-noise ratio

When a minimum usable output signal-to-noise ratio is specified, Eq. (2) can be interpreted as an expression for receiver sensitivity. Operating noise temperature as defined here is directly related to operating noise figure as used by P. D. Strum and others. The relation is simply

$$T_{op} = F_{op} T_o \quad \text{where} \qquad (3)$$

F_{op} = operating noise figure

For amplifiers other than the degenerate parametric amplifier it is easily shown that

$$T_{op} = (F-1)\,T_o + T_A \quad \text{where} \qquad (4)$$

T_A = antenna temperature

For the degenerate parametric amplifier which is receiving normal coherent signals (i.e., not noise), the corresponding operating noise temperature is

$$(T_{op})_d = 2\left[(F_{dsb}-1)\,T_o + T_A\right] \qquad (5)$$

The 3.5 db traveling-wave tube will thus have an operating noise temperature of $360^{\circ}K + T_A$ while the 2 db parametric amplifier will have an operating noise temperature of $340^{\circ}K + 2T_A$. Therefore, when the antenna temperature is greater than $20^{\circ}K$, the traveling-wave amplifier will have the better noise performance.

There are two special cases where the above method of comparison is not applicable. One is the case when synchronous detection is used, and the other is when broadband noise is the signal, as in radiometry. According to Adler and others, in both of these cases the double-sideband noise figure is directly applicable. In radiometry, however, the whole story is not told when the noise figure is specified, for the rf bandwidth is also of importance. A proper figure of merit which is proportional to the minimum detectable temperature difference would be the ratio of the operating noise temperature to the square root of the instantaneous rf bandwidth. Continuing our example, let us suppose that the 2 db parametric amplifier has a bandwidth of 200 mc, while the 3.5 db traveling-wave tube has a bandwidth of 2,000 mc. Assuming room temperature for the antenna temperature and remembering that in this case the double-sideband noise figure is appropriate, we see that the parametric amplifier will have a figure of merit of

$$\text{Figure of Merit} = \frac{1.58}{\sqrt{200}} = 0.112$$

The traveling-wave tube will have a corresponding figure of merit of

$$\text{Figure of Merit} = \frac{2.24}{\sqrt{2,000}} = 0.050$$

Here again we see that the traveling-wave tube would be superior to the parametric amplifier, since it would be capable of resolving a smaller difference in temperature.

These examples are not intended to imply that the traveling-wave tube is "superior" to the parametric amplifier. Rather, they are intended to indicate that because a parametric amplifier has a "lower noise figure," it does not necessarily follow that it will exhibit better performance in an actual system. In many instances it will, but as has been indicated in the above examples, there are some applications where it may not.

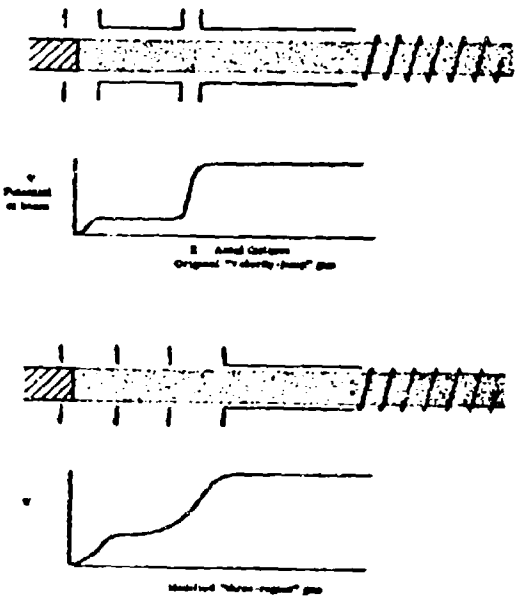


Fig. 1 - Early electron gun designs for use in low-noise traveling-wave tubes. The first of these, the "velocity-jump" gun, featured a rather abrupt change or "jump" in the potential along the beam. In a subsequent design the control of potential along the axis was divided into three regions by the use of auxiliary anodes. When optimized, the potential profile of the three-region gun exhibited a smoother transition from cathode to rf circuit than in the earlier "velocity-jump" gun.

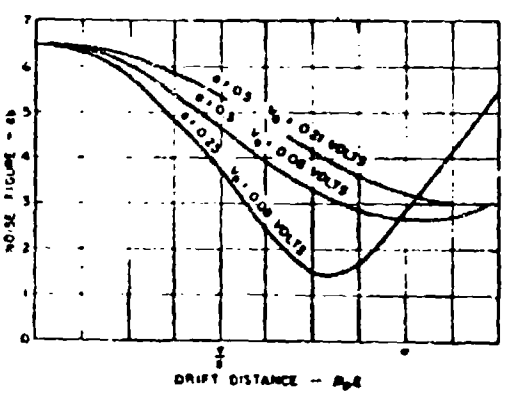
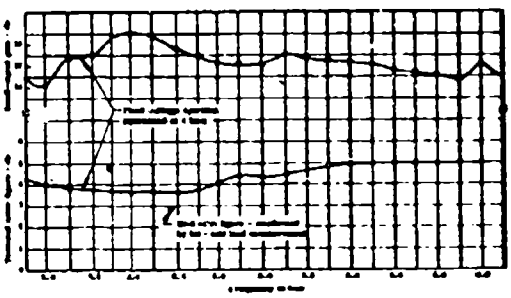


Fig. 2 - Results of a theoretical study in which the noise content of an electron beam was computed as a function of distance in a very low velocity drifting region. The predicted noise figure of an amplifier using these beam values is plotted as a function of the drift length with beam current density as the essential parameter. Noise figures in the 1 to 2 db region are indicated as possible.



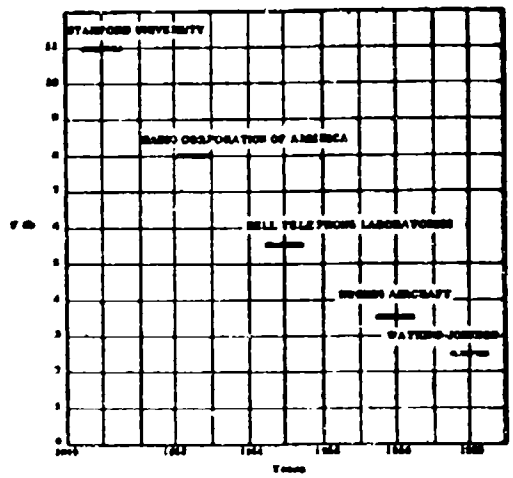


Fig. 5 - A plot of minimum noise figures obtained over the past eleven years. Significant developments have resulted in a plateau-like quality to the graph.

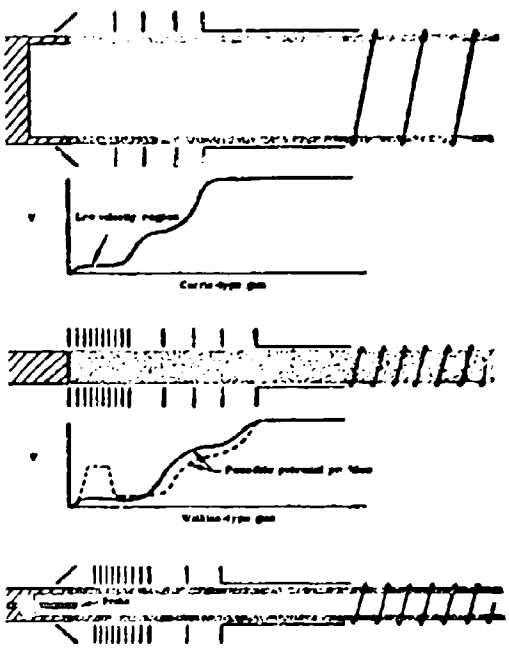


Fig. 6 - Several electron gun configurations which have been used in obtaining continued noise improvement. The principal feature in all is more precise control of low-velocity regions within the gun. Various combinations of these gun elements are also possible. It is postulated that continued studies of these and other guns will lead to traveling-wave noise figures of 1 to 3 db.

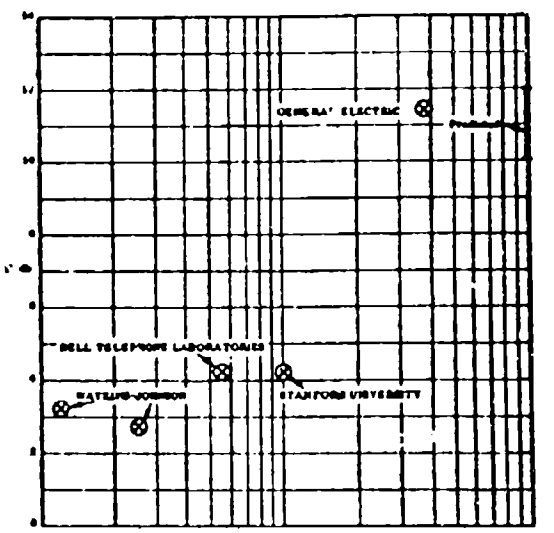


Fig. 7 - A plot of minimum noise figures measured in traveling-wave tubes as a function of frequency. The small spread in values over the 1 to 1 ratio from 1000 to 10,000 mcs encourages extrapolation to higher frequencies and 10 to 12 db is predicted as obtainable at 100,000 mcs.

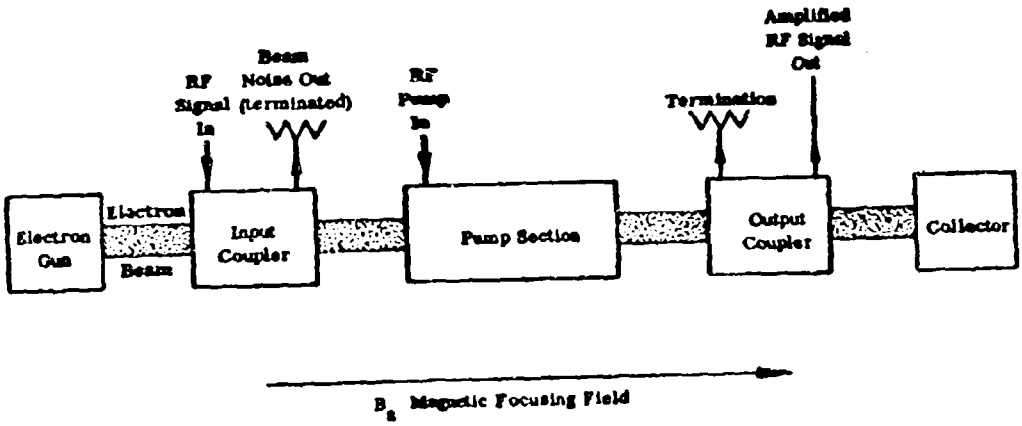


Fig. 7 - A schematic diagram of the essential elements of a beam-type parametric amplifier. The use of a discrete electron beam results in physical separation of the functional elements. The input coupler serves the dual role of transferring the rf signal to the beam while coupling the fast-wave noise out of the beam prior to amplification. Signal amplification occurs in the pump region where rf energy is supplied from an external source. The output coupler serves to remove the amplified signal from the beam.

BULK SEMICONDUCTOR DEVICES FOR MICROWAVE APPLICATIONS

By: H. Jacobs*, F. A. Brand*, J. Meindl*, M. Bennett* and B. Benjamin**

SUMMARY

In considering the transmission of microwave energy through a semiconductor medium such as high resistivity germanium, an analysis has been made of the effects of absorption and multiple reflections within the medium. By varying the conductivity, changes in the power transmitted have been observed. New device possibilities appear feasible such as an amplitude modulator with little or no change in phase shift, a phase modulator with no change in amplitude, and other optical-electrical devices such as low frequency amplification and beam directional devices.

INTRODUCTION

It has been observed recently that when a semiconductor rod is inserted in a waveguide, modulation of the transmitted power can be obtained by varying the conductivity of the semiconductor.[1] In what follows, various theoretical and experimental considerations lead to evidence that a family of new type devices may be attainable such as an amplitude modulator, a phase shifter, and other microwave components.

We start first by considering the transmission of microwave energy simultaneously through three media, first air then a semiconductor slab, infinite in length in the x and y directions but finite in the z direction and third in air again. The proposed scheme is illustrated in Figure 1. In such an arrangement one would expect transmission through the semiconductor slab to be dependent upon absorption in the semiconductor and possible multiple reflections. Hence the transmission would be found to be both a function of the conductivity and the thickness of the slab. This type problem has been discussed by Stratton[2] and Bromwell and Beam[3] and is found to be most easily approached by use of the transmission line analogue of a wave travelling in free space. In Figure 1 the transmission line is indicated in its relation to the electromagnetic propagation systems. In order to calculate the role of changes of conductivity and thickness on the changes in transmission of the electric field through the medium the following relations were used,

$$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)}, \qquad (1)$$

$$\eta = \frac{j\omega\mu}{\gamma}, \qquad (2)$$

$$z = \eta_2 \frac{\eta_3 + \eta_2 \tanh \gamma_2 l_2}{\eta_2 + \eta_3 \tanh \gamma_2 l_2}, \qquad (3)$$

$$r_t = \frac{2z}{z + \eta_1}$$

*U. S. Army Signal Research and Development Laboratory, Fort Monmouth, New Jersey
**Monmouth College, West Long Branch, New Jersey

and $\frac{E_2}{E_{1a}} = r_t \left(\cosh \gamma_2 l_2 - \frac{\eta_2}{z} \sinh \gamma_2 l_2\right)$. (4)

In these relations, the following values for the various constants were assumed,

ω = 10,000 megacycles,

ϵ = 16 x 8.854 x 10^{-12} farads/meter,

μ = 1.257 x 10^{-6} henry/meter,

σ = 2, 3, 4, 6 and 10 (ohm meters)$^{-1}$,

z = input impedance at the plane ab,

$\eta_1 = \eta_3$ = 376.6 ohms,

η_2 = intrinsic impedance of the semiconductor using (1) and (2) with various values of σ,

γ_2 = propagation constant of semiconductor using (1) for various values of σ,

$\gamma_1 = \gamma_3 = j\beta_0, \beta = 2\pi/\lambda_0$

l_2 = thickness of the slab,

r_t = ratio of electric intensity in semiconductor medium to the electric intensity of incident wave in air at the front surface,

E_0 = the electric field transmitted through medium 2 into region 3,

E_{1a} = the electric field in medium 1 incident on the front surface.

In considering (1) and (2), these can further be broken down to

$$\gamma_2 = \alpha_2 + j\beta_2 \qquad (5)$$

and

$$\eta_2 = x_\eta + jy_\eta. \qquad (6)$$

The specific values are listed in Table I for the propagation characteristics of electromagnetic waves in germanium at 10,000 megacycles per second. Using these data the following information is shown as a function of conductivity of the germanium and the thickness, l_2, E_0/E_{1a} the magnitude of the ratio of the transmitted electric intensity to the electric intensity in air incident upon the surface, the phase angles, in radians, of E_0 with respect to E_{1a}, the magnitude of r_t and the phase angles of r_t, in Figures 2, 3, 4 and 5 respectively. In considering these data we can postulate the existence of device possibilities for amplitude and phase modulation. In explaining Figure 2, if we assume a constant thickness l_2, we see that E_0/E_{1a} will vary with conductivity. For instance at $\frac{1}{4}$ millimeters thickness, varying the conductivity of the slab

from $\sigma' = 2$ to $\sigma' = 10$, by some physical means such as light or uniform injection of excess minority carriers, will decrease E_o/E_{in} from roughly 56 percent to 10 percent. Next we consider Figure 3 for the phase shift to be expected due to conductivity modulation of the electric intensity. We note that at about 2.1 millimeters thickness a node exists due to the internal reflections. In other terms, if the conductivity is modulated with this particular thickness, amplitude modulation is attained without phase shift. This is done to some extent with a decrease in the amount of amplitude modulation. However, the attractive possibility does exist for a significant amplitude change with no phase shift.

Suppose now one would like to attain phase shift without amplitude modulation. This does not appear as quantitatively but can be shown to be possible. If we choose a six millimeter thickness, considerable phase shift appears in Figure 3. Referring again to Figure 2, we see a valley at the same thickness. Now this valley occurs due to multiple internal reflections which in turn is dependent upon uniformity of thickness and conductivity. If excess carriers could be injected in the germanium so as to be slightly non uniform, the internal reflections would be destroyed and although the phase shift would still be changed by the excess carriers, the transmission amplitude will not necessarily change and under specific circumstances may even rise.

With this theoretical introduction we shall next consider design principles for various devices together with experiments performed to illustrate their feasibility.

AMPLITUDE MODULATOR

So far we have considered the ideal situation of an infinite slab in the free space oriented perpendicular to the direction of propagation of the electromagnetic wave. We can extend this concept to where the semiconductor slab is inserted into a waveguide perpendicular to the flow of energy[4] or in parallel with the flow. The two possible orientations are illustrated in Figure 6. In addition, contacts are placed on the sides of the semiconductor for the purpose of injecting excess minority carriers into the semiconductor medium by electrical means. If the sample is oriented as shown in Figure 6a we should expect to observe the multiple internal reflection phenomena described above. However, if the sample is oriented as in Figure 6b, the situation is much more complicated. In the latter arrangement the internal reflections have been found to be smaller and the absorption by conductivity changes of the semiconductor is the more pronounced factor in causing modulation of the transmitted signal. If we make certain simplifying assumptions we can design the geometry of the semiconductor shown in Figure 6b to provide the maximum changes in power for a given change in conductivity.

We assume first that the bulk lifetime of minority carriers in the sample is long, but still of such magnitude that it is not affected by the changes in dimensions to be discussed. We assume further that in this orientation, the changes in transmission are mainly caused by absorption as excess minority carriers are injected. The third assumption is that the power attenuation factor, α, increases linearly with conductivity, σ'. The last assumption is approximately true if $\sigma' \ll \omega\epsilon$. With these assumptions, we can carry out calculations giving the first order magnitude of dimensions and effects to be expected. It must be realized that these data will have to be later modified if one considers multiple reflections. Hence, only as a starting point, we consider an infinite medium of germanium in which the propagation constants are related to frequency, conductivity and dielectric constant.

For a uniform, infinite medium, the absorption of power is given by,

$$P = P_o e^{-\alpha x}, \tag{7}$$

where P_o is the incident power, P the power at the distance x, and α is the attenuation constant. The change in power with respect to α is

$$\frac{dP}{d\alpha} = -x P_o e^{-\alpha x}. \tag{8}$$

By differentiating (8) and setting the result equal to zero, one can find that $dP/d\alpha$ is a maximum at the thickness given by

$$x = \frac{1}{\alpha}. \tag{9}$$

In the form α, the power attenuation factor is equal to $2\alpha_o$, where α_o is the attenuation factor for the electric field. In Table II, numerical values of α_o are given as calculated for germanium and electromagnetic radiation at 10,000 megacycles. In addition, the values of optimum length are listed for each specific conductivity.

Experimental samples were made up using germanium with conductivity in the order of magnitude of 3(ohm meters)$^{-1}$ and in the geometry shown in Figure 6b. The width of the sample was uniformly .25 cms, the height fixed at .8 cms and the lengths were varied from .198 cms to .8 cms. At a frequency of 9600 megacycles per second and in an X band waveguide, the indications are that the maximum sensitivity, $dP/d\sigma'$, occurs in the region of .35 cms or slightly higher.

In Figures 7, 8, 9 and 10 we see the experimental results of injecting excess minority carriers in samples with dimensions and orientations as described. Here we have power attenuation as a function of current injected into the semiconductor. Now since α is assumed linearly related to σ', the maximum $dP/d\alpha$ will occur at the same length as the maximum $dP/d\sigma'$. In these tests we are measuring current and hence for the same current we are producing a different change in conductivity if the volume is different. Therefore the experimental data can only be used to verify the theory in an approximate manner.

If we assume at the higher currents the sample is flooded with injected excess minority carriers,

and that the current is a measure of the average number of excess minority carriers, we should expect that for a given high current, the attenuation would be greatest at about .35 cms length. In examining Figures 7 through 10 we see that this is approximately the case, even if the volume is taken into account. For instance, for 8 millimeter length at 50 ma, the attenuation is 3.8 db. At 6 millimeter length, the volume is .75 that of the first case and we consider .75 x 50, or 37.5 ma for equivalent density of excess carriers in Figure 9. Here the attenuation is about 4.5 db. At 3.5 millimeter length the volume is .436 that of the 8 millimeter sample. The current to be considered in Figure 8 is .436 x 50, or 21.8 ma. The attenuation here is 4.3 db. With reference to Figure 7, the volume is decreased to .24 of the 8 millimeter sample and the current is .24 x 50 or 12.0 ma. At 12 ma the attenuation is 1.95 db. Hence even in normalising the volume, the maximum change attenuation occurs at high currents in the range from 3.5 to 6.0 millimeter length.

For small injection currents i.e., under 5 ma, (7), (8) and (9) may not be applicable due to the fact that under these conditions the carriers are localised in a small spot near the injecting junction. Under these conditions the volume would not effect the amount of attenuation as much as the total number of injected carriers in the localised area.

In order to calculate the insertion loss for the arrangement in Figure 6b, we again start with (7) and the loss in the sample (assuming reflection changes are small) is given by,

$$1 - \frac{P}{P_o} = 1 - e^{-\alpha x}. \qquad (10)$$

Consider X the thickness for maximum $dP/d\alpha$. Since $X = 1/\alpha$ we have,

$$1 - \frac{P}{P_o} = 1 - e^{-1}$$

$$1 - \frac{P}{P_o} = 63.2\%. \qquad (11)$$

In general, as thickness increases, the insertion loss increases. Another term which has been found useful is the ratio $\Delta P/P$. Here P is the power transmitted past the semiconductor block and arriving at a tuned detector, when no current is injected into the germanium rod. As current is injected, the decrease in power reaching the detector is defined as ΔP. In a sense the ratio $\Delta P/P$ can be defined as a modulation ratio. Again starting with (7)

$$\frac{P_2 - P_1}{P_2} = \frac{P_o\left(e^{-\alpha_2 x} - e^{-\alpha_1 x}\right)}{P_o\, e^{-\alpha_2 x}}, \qquad (12)$$

where α_2 and P_2 correspond to attenuation factor and power reaching the detector with no current injection into the sample and α_1 and P_1 indicate the attenuation factor and the power arriving at the detector with current injected into the sample. In practice, $\alpha_2 < \alpha_1$ and $P_2 > P_1$. By (12),

$$\frac{\Delta P}{P} = 1 - e^{-(\alpha_1 - \alpha_2)x} \qquad (13)$$

If X is now chosen as $1/\alpha_2$, the thickness for maximum $dP/d\alpha$,

$$\frac{\Delta P}{P} = 1 - e^{-\alpha_1/\alpha_2 + 1}. \qquad (14)$$

We note that since $\alpha_1 > \alpha_2$, $\Delta P/P$ will increase with increasing α_1 monotonically. Conversely, if $\alpha_1 \to \alpha_2$, $\Delta P/P \to 0$. One further point to be considered is the phase shift which may occur during the attenuation of the wave due to the increased conductivity of the germanium rod.

In general, one would expect some phase shift due to a change in conductivity. In considering Figure 11 calculated for an infinite medium of germanium, we see that at 10,000 megacycles there is only a very slight shift in wavelength with conductivity in the low conductivity region. It would seem that by the correct adjustment of geometry, or by decreasing the length of the sample in the direction of energy flow, one could obtain the dimensions of the rod such that little or no phase shift would occur. Just such circumstances have been found. In Figure 12 we see data indicating amplitude modulation with no phase shift.

PHASE SHIFTING DEVICES

Previously we have discussed devices which will give amplitude modulation with no phase shift. Now let us consider devices aimed at providing phase shift with little or no amplitude modulation. As was described in the introduction, if the sample is oriented to take into account internal multiple reflections (Figure 6a), the E_t/E_{in} magnitude may lie in a valley at various thicknesses of the sample. If current is injected not entirely uniformly, the phase can be shifted due to increases in conductivity, and the absorption component increased, but due to the destruction of the uniformity of internal reflections, the actual transmission magnitude may stay the same or even increase with increasing conductivity of the germanium slab. Experiments have been conducted in which these effects have been noted. For example, one sample was formed with dimensions of .8 x .8 x 1.5 cms and then inserted in the waveguide such that the large dimension was across the guide as indicated in Figure 6a. When tested in the microwave bridge approximately 30° phase shift was noted with small changes in attenuation. The data here is indicated in Table III. This experiment was further checked by flooding the sample with light to increase its conductivity. Again it was verified that increased conductivity (induced by the light source) could actually cause an increase in transmission at the same time the phase shift was changing.

OTHER DEVICE POSSIBILITIES

Still other devices are implied in these physical effects. For instance, if the PIN diode is reverse biased, conductivity modulation can occur at higher frequencies than when biased in the forward direction. The design for an amplifier has been worked out using this principle.

Finally, since these principles apply to the infrared as well as the microwave region, various infrared modulators, phase shifters and detectors could be possibly arranged to carry out new electro-optical functions.

ACKNOWLEDGMENT

The authors are indebted to Mr. L. Mathis, Dr. G. Cochran, Miss E. Ulrich and Dr. W. Gaertner for the help they have rendered in the form of many discussions during the course of this work and to Dr. W. Caldwell of the Bendix Aviation Corporation, Long Branch, New Jersey for furnishing germanium samples.

REFERENCES

1) A. F. Gibson, "Progress in Semiconductors", John Wiley & Sons, Inc., New York, N. Y., Vol. 2, p. 229; 1957.

   A. P. Ramss, H. Jacobs and F. A. Brand, "Microwave Techniques in Measurement of Lifetime in Germanium", J. Appl. Phys., Vol. 30, pp. 1054-1060; July 1959.

   W. J. Harrick, "Metal to Semiconductor Contacts: Injection or Extraction for Either Direction of Current Flow", Phys. Rev., Vol. 115, pp. 876-882; 15 August 1959.

2) J. A. Stratton, "Electromagnetic Theory", McGraw-Hill Book Company, Inc., New York, N. Y., p. 515; 1941.

3) A. B. Bronwell and R. E. Beam, "Theory and Application of Microwaves", McGraw-Hill Book Company, Inc., New York, N. Y., p. 279; 1947.

4) If a slab of high resistivity germanium is arranged to fill a section of the waveguide, it turns out that in the experiment described, the guide wavelength is not very different from the wavelength in an infinite slab of germanium. For instance, in the case of a plane wave at 10,000 megacycles the wavelength is approximately .75 cms in germanium. For the case of the semiconductor filling the waveguide and $\lambda_0 = 4.65$, the guide wavelength is approximately .76 cms.

TABLE I

PROPAGATION CONSTANTS IN GERMANIUM AT 10,000 MEGACYCLES PER SECOND

σ (ohm meters)$^{-1}$	γ_2	η_2
2	$848.0\, e^{j1.460}$	$93.04\, e^{j.1105}$
3	$887.7\, e^{j1.409}$	$92.98\, e^{j.1447}$
4	$905.6\, e^{j1.357}$	$89.93\, e^{j.1788}$
6	$945.1\, e^{j1.284}$	$84.80\, e^{j.2484}$
10	$1028.0\, e^{j1.140}$	$76.81\, e^{j.4217}$

TABLE II

ATTENUATION CONSTANT AS A FUNCTION OF CONDUCTIVITY

σ (ohm meters)$^{-1}$	α_0 nepers/meter	$\frac{1}{2\alpha_0} = X$ optimum length in meters
2	93	5.4×10^{-3}
3	143	3.5×10^{-3}
4	192	2.6×10^{-3}
6	270	1.9×10^{-3}
10	420	1.2×10^{-3}

TABLE III

DATA FOR PHASE SHIFT DEVICE WITH SMALL ATTENUATION CHANGE AT 9600 MEGACYCLES PER SECOND

Change in attenuation in db with respect to no current injected	Phase Shift in degrees	Current injected in milliamperes
+.31	8.6	2
+.46	13.5	4
+.37	16.5	6
+.35	20.2	8
+.26	22.2	10
+.07	25.7	15
-.08	27.2	20
-.18	29.7	25
-.44	31.7	30

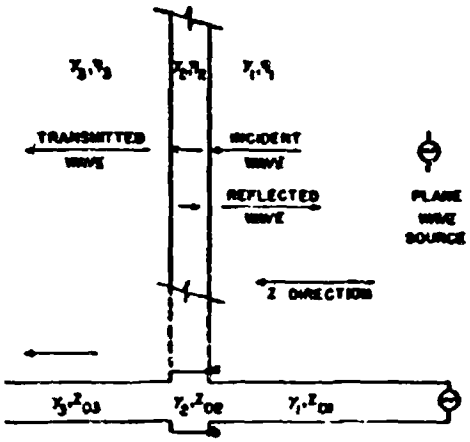


Figure 1

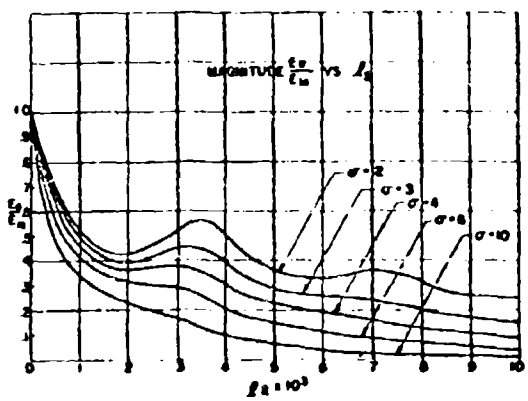


Figure 2

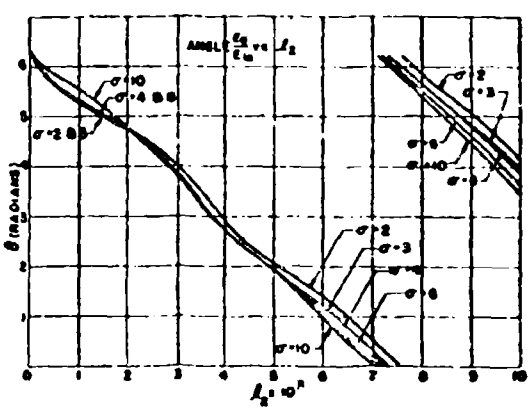


Figure 3

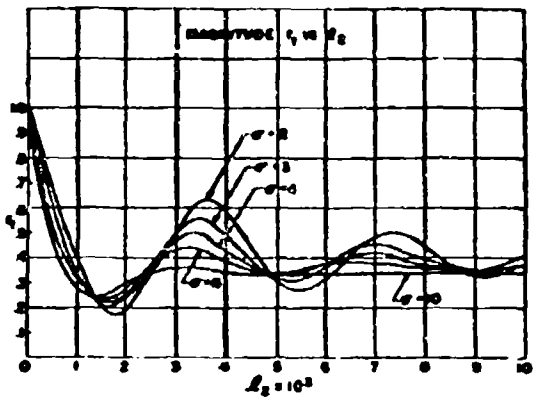


Figure 4

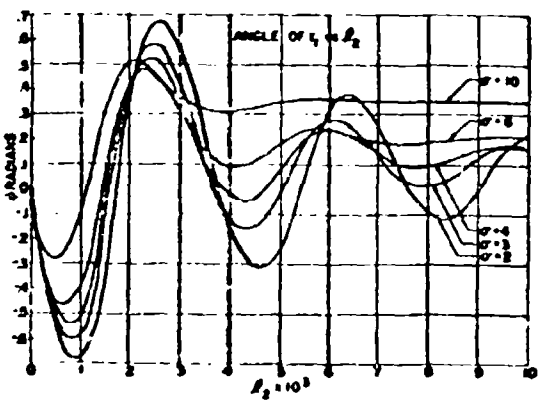
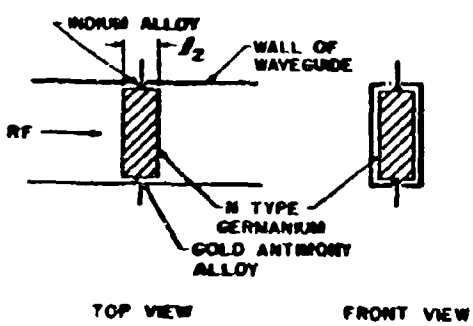
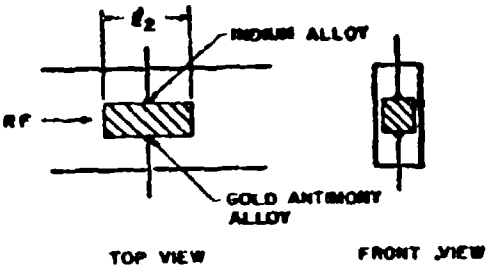


Figure 5



6-a

Figure 6A



6-b

Figure 6B

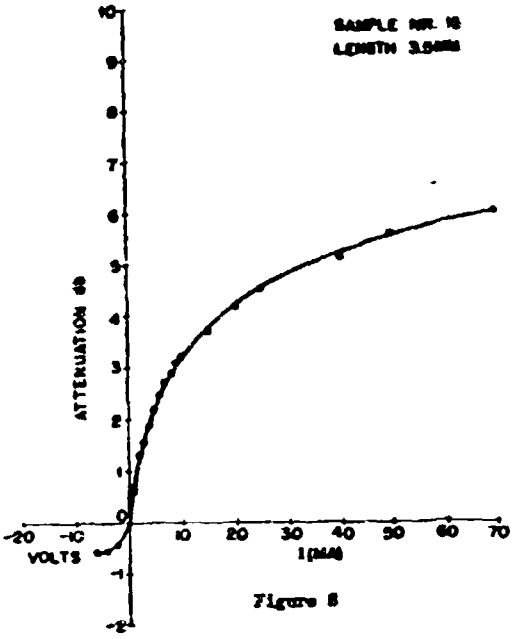


Figure 8

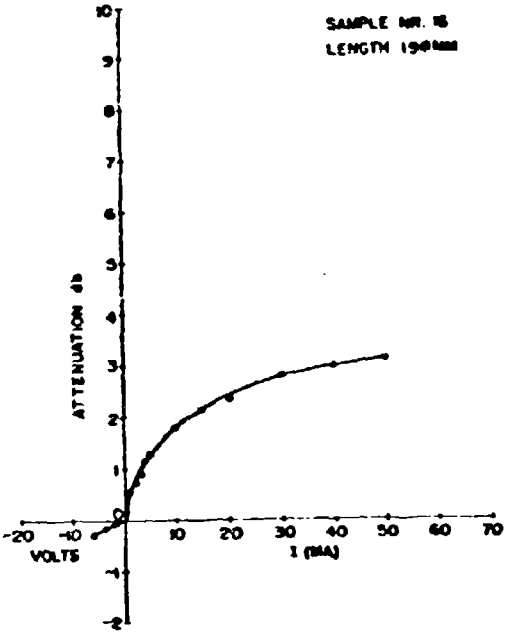


Figure 7

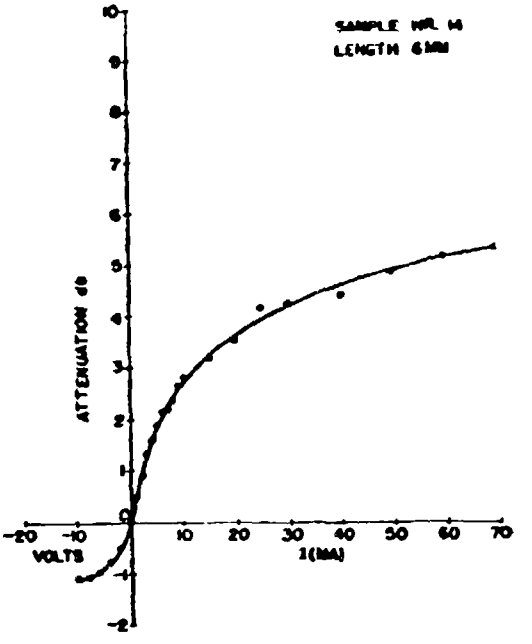


Figure 9

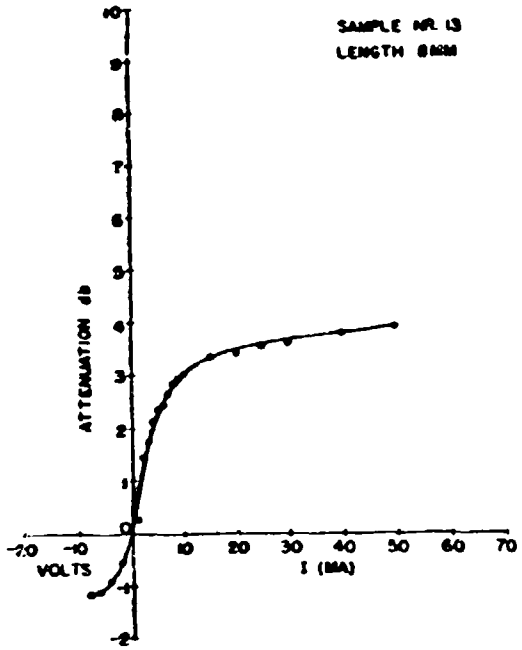


Figure 10

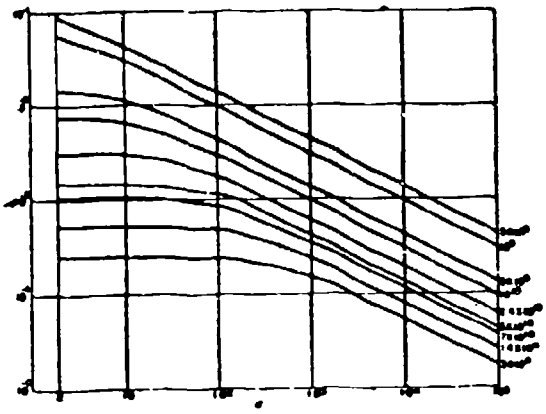


Figure 11

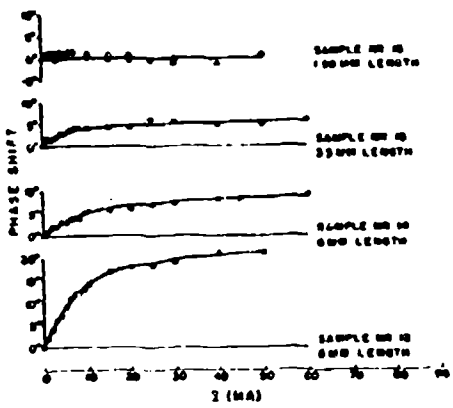


Figure 12

LIST OF CAPTIONS

Figure 1: Transmission line analogue for a plane wave transmitted through a slab of dielectric. In this case, medium 1 and medium 3 are air and medium 2 is germanium of low conductivity.

Figure 2: Magnitude of the ratio of electric field intensity, E_{2s}, transmitted through the semiconductor to the electric field incident, E_{1s}, upon the front surface as a function of thickness and conductivity.

Figure 3: The phase angle of the transmitted electric field, E_{2s}, with respect to the electric field incident on the front surface, E_{1s}, as a function of thickness and conductivity.

Figure 4: The magnitude of r_t, the ratio of the electric field just under the front surface of the semiconductor to the incident field in air as a function of thickness and conductivity.

Figure 5: The phase angle of r_t as a function of thickness and conductivity.

Figure 6: Two arrangements for orientation of a semiconductor slab in a waveguide. In Figure 6a, the largest phase is oriented perpendicular to the flow of microwave power. In Figure 6b, the same plane is oriented parallel to the power flow. The material is 39 ohm cm, N type germanium and the indium, germanium, gold antimony structure forms a type of PIN diode.

Figure 7: Attenuation of power as a function of current injected for a sample with l_2 = .196 cms. The orientation is described in Figure 6b. The negative attenuation refers to increased transmission as PIN diode is biased in the reverse direction. Frequency was 9600 megacycles per second.

Figure 8: Attenuation of microwave power as a function of forward current and reverse voltage for a sample length, l_2 = .35 cms. Frequency was 9600 megacycles per second.

Figure 9: Attenuation of microwave power as a function of forward current and reverse voltage for a sample length, l_2 = .60 cms. Frequency was 9600 megacycles per second.

Figure 10: Attenuation of microwave power as a function of forward current and reverse voltage for a sample length, l_2 = .80 cms. Frequency was 9600 megacycles per second.

Figure 11: Wavelength, λ, of electromagnetic waves in a medium of germanium as a function of conductivity σ. The frequency is indicated along each solid line.

Figure 12: The phase shift of samples oriented as shown in Figure 6b, with l_2 varying from .193 cms to .80 cms. The phase shift was measured concurrently with attenuation in a microwave bridge as the forward current was injected in the PIN diodes. At approximately a length of .2 cms the phase shift becomes negligible although over 3.0 db of attenuation was obtained.

HIT INDICATOR TECHNIQUES FOR DIRECT FIRE WEAPONS

By: Herbert Chaskin, Project Engineer, U. S. Naval Training Device Center

Introduction

A hit indicator system for direct fire weapons has recently been developed to offer improved training for Army personnel competing in land combat under a battlefield environment. Special techniques are used to simulate the effects of properly firing a tactical weapon at personnel, vehicle and fortification targets. The basic system can best be described by Figure 1 which illustrates simulator equipment offering weapon and target capabilities for an infantry rifleman in the field.

The weapon equipment consists of:

(a) A photoelectric telescope

(b) A firer's transmitter

(c) A trigger switch mounted within the weapon which activates the system when the trigger is pulled.

The target equipment consists of:

(d) A target receiver

(e) An omnidirectional xenon flash tube assembly

(f) A common antenna for both equipments.

The fundamental problems of the system are divided into the following three categories:

1. Identification of the intended target with respect to other targets.

2. Proper superelevation of the weapon to correlate with target range.

3. Confining the sensitive field of view of the photoelectric telescope to the area of a target changing in size as a function of range.

Description

A block diagram of the system in Figure 2A illustrates the operation of the basic firer equipment. Pre-travel of the trigger of the weapon initiates a microswitch, which closes a relay within the firer's pack. This energizes all transistor circuits and initiates power to a 150 Megacycle F.M. transmitter. Proper trigger squeeze establishes sufficient warm up time for the filaments of the transmitter. When the weapon is fired, the vibration of the sear releasing the hammer is picked up by a barium titanate crystal. This signal is amplified and transferred to a coding circuit which modulates an F.M. transmitter and initiates an RF interrogation pulse. The interrogation pulse is transmitted omnidirectionally to all targets within the effective range of the weapon. A block diagram of the system in Figure 2b illustrates the operation of the basic target equipment. The problem of target identification is resolved by coding the time at which the target flashes with respect to the time of the interrogation pulse. Upon receipt of the interrogation pulse, the target starts a code generator which produces two pulses positioned in time from the interrogation pulse. This coding circuit is coupled to a flash tube trigger circuit which flashes each lamp from a 450 volt power supply. Each target has a unique combination of coded intervals, thereby applying a signature of the target to its coded flashes. Detection of a flash from a specific target results only when the weapon is properly aimed and fired. The photoelectric telescope of the weapon detects these flashes and transmits a one millisecond R.F. pulse coincident with each flash to all targets. Through use of decoding circuits, each target compares the receipt of both R.F. pulses with the first and second coded delay times of each flash. Only the target that receives R.F. pulses coincident with both flashes is capable of indicating a hit. The target, now in a state of potential hit, waits to determine its vulnerability to the weapon that fired. Each weapon is coded so that a single R.F. pulse is transmitted when a small arms weapon is fired and two pulses for larger weapons. If vulnerability is established, then the hit indicator system at the target generates a 3000 cycle tone at the speaker for a ten-second interval. At the end of this period, a pyrotechnic device is electrically triggered, producing a flash and smoke indication at the target. An electrical fuse is also blown which deactivates all weapon and target equipment.

Sequence of Events

A time sequence diagram in Figure 3 illustrates the time relationships between the operation of weapon and target equipments. The first line represents trigger action. Transmitter warm-up occurs when the trigger is pulled, and in its pre-travel prior to releasing the hammer. When the hammer is released, the vibration detector that is part of the trigger switch assembly picks up a vibration shown on the second line. The first vibration is that of hammer release and occurs at t_o. This is of relatively small amplitude as compared to the firing of a blank round 8 to 10 milliseconds afterwards.

The next line illustrates transmitted pulses from weapon to target occuring at t_o and coincidently with each target flash. The R.F. pulse transmitted at t_o + 10 milliseconds results from the detection of the blank round being fired for a small arms weapon. A second pulse occurs at t_o + 11 milliseconds for larger caliber weapons.

Below the dotted line is a diagram of the timing sequences of coding and decoding circuits at the target. The first code interval for this

particular target is initiated at t_0 and is 3 milliseconds long. This pulse causes flash tube No. 1 to trigger at $t_0 + 3$ milliseconds. The flashing of this lamp triggers a second code interval 4 milliseconds in duration. At the end of $t_0 + 7$ milliseconds, a second flash tube is triggered. The hit indicator device is activated after all coded conditions are satisfied.

Grey Scale Reticle

Problems of correlating target range with weapon superelevation and confining the sensitive field of view of the photoelectric telescope to the target area is resolved by a specially designed grey scale reticle illustrated in Figure 4. The grey scale reticle is capable of measuring an amplitude of signal correlated with range by comparing this amplitude with the angle through which the signal is received. This amounts to the rifleman's estimate of range in terms of superelevation of the weapon. The device also serves to automatically change the field of view of the photoelectric telescope with changing range. This provides a vulnerable target area of constant size that is independent of the target range from the weapon.

The field of view of the telescope is restricted by mounting a field stop at the focal length of the objective which is in the first image plane of the optical system. This field stop is completely opaque except for a hole in its center. The shape of this hole defines the shape of the projected field of view. If this field stop were triangular in shape, and designed to be projected at some specific range so that it was 3 feet high and 3 feet across at its base, it would project a sensitive area, triangular in shape as shown in Figure A. The point of aim, that is the optical axis of the telescope is bore sighted with the line of sight on the weapon, and is adjusted so that it falls at the vertex of this triangle. If the weapon is then aimed at the light source on a personnel target, infrared emitted from this source will pass through the field stop and will be detected. This inverted triangle can be displaced downward and to the left as in Figure (b) so that the point of aim is at the left hip of the man. Infrared could still be emitted through the field stop and detected. The same applies to the right side of the man as indicated in Figure (c). This then results in a vulnerable area shown in Figure (d), the shape of which is an inversion to the projected sensitive area. If an aim point is selected anywhere within this triangular vulnerable area a hit will be detected. The problem now is to make the size of this triangle change automatically with range. This is possible due to the requirement that the rifleman elevate his weapon to compensate for projectile trajectory with increasing range. This results in effectively sliding the desired vulnerable area as shown in Figure (e) up or down as range is increased or decreased. This vulnerable area shown in Figure d must become smaller with increasing range. The result is a reticle with curved sides and a flat base three feet across at the minimum range of the weapon. This varies in width as the three foot dimension subtends a smaller and smaller angle with increasing range. A target at maximum range would require a vulnerable area consisting of only a small portion of the vertex of this entire vulnerable area. As the target moves closer in range, this triangular pattern moves down across the vulnerable area and steadily increases in height maintaining a 3 foot high area until the minimum range point is reached. This results in the formation of a trapezoidal area of vulnerability instead of the triangular area shown in Figure (d).

The problem is now to establish an upper and lower base line for this trapezoid. This is accomplished by making the reticle field stop of the telescope a grey scale. This scale is calibrated in density in accordance with the firing tables of the weapon which show the required elevation of the weapon at all ranges. The grey scale varies in density from clear at the maximum range end of the projected area to a very dense area at the minimum range end of the pattern. This then means that the signal received by the phototransistor will be of equal amplitude for all ranges if the weapon is programmed properly in superelevation as the target changes in range.

The output of the phototransistor feeds a miss detecting circuit located in the firer's unit which requires a minimum threshold signal to trigger the hit indicating circuit. This circuit is provided with an upper cutoff limit such that if the weapon is aimed too high, the signal received would be in excess to that required for the range of the target. A cutoff circuit is triggered which prevents the indication of a hit. Figure (e) shows a condition wherein the weapon was aimed too low. This results in the image of the flash tube passing through a dense part of the grey scale and thereby being attenuated to a point below the minimum threshold detecting level.

Figure (f) indicates a situation where the weapon was aimed too high. The clear portion of the scale was used resulting in a signal of sufficient amplitude to trigger the cutoff tube circuit resulting in a miss.

Figure (g) depicts a situation where the image falls through the vulnerable portion of grey scale providing proper superelevation was estimated and under this condition a hit occurs.

An exploded view of the photoelectric telescope is given in Figure (5). This unit is machined so that its mounting is boresighted to the barrel of the rifle. The telescope uses f_2 objective lenses, 26mm in diameter. These lenses focus the image on the reticle field stop shown immediately behind the main barrel of the telescope. Signal passing through the field stop

is collected by a second group of optics and focused on a small collector lens. This lens focuses the signal on a G.E.ITE 1393 phototransistor. An amplifier within the telescope has been designed to provide both gain and D.C. Stabilization for the detector. A filter mounted between the second lens group and the phototransistor, eliminates the possibility of saturating the detector with visible light.

Flash Lamp Assembly

The flash lamp assembly shown in Figure 6 consists of two General Electric xenon flash lamps mounted to a common base. These lamps are loaded to approximately 20 watt seconds which extends its life to over 100,000 flashes. The lamps are supplied by storage capacitors charged to 450 volts and require a trigger potential of 10 volts at 20 millijoules. Adjacent to the flash lamps shown in Figure 6 is an infrared filter. Little or no visible light can be detected radiating from these flash lamps because a filter is used. A lucite dome is used to provide mechanical protection to the lamps. The shield shown in Figure 6 adjacent to the plastic dome is used on the flash lamp associated with vehicle and fortification targets. This shield limits the radiating field of the flash lamp required for vehicle and installation targets. In the case of the personnel target, the flash lamp assembly is mounted on the helmet and does not have this shield. Also mounted to the helmet and potted in the base of the flash lamp assembly are high voltage trigger coils and an R.F. antenna connection.

M60 Equipment

M60 Machine Gun Equipment is illustrated in Figure 7. The operation of the hit indicator system requires some additional circuits to simulate a burst of fire since the M60 Machine Gun is not capable of automatic operation with blank rounds. These circuits are energized by loading a single blank round and firing it. The vibration detected, triggers a circuit that free-runs up to a maximum of six pulses per blank round. The telescope used with the M60 is designed to have a maximum range of 1000 yards. The grey scale pattern for the M60 machine gun is similar to that of the M14 rifle with the exception that it is much higher in size due to the greater superelevation required of the weapon at maximum range.

106mm Recoilless Rifle

The basic system has also been developed for a 106mm recoilless rifle operating over a range of 2000 yards. A photoelectric telescope of increased sensitivity using a photomultiplier is required. Normal aiming and firing techniques are simulated such that a target image is viewed on a reticle pattern of the weapon sight. Errors in range estimation, normally demonstrated by a 50 caliber spotting round are indicated as high or low misses within a sight adapter. This high or low miss detection is possible by using the grey scale principle described earlier and by increased optical sensitivity. Signals passing through the dense portion of the grey scale normally below the threshold level of a hit indication, are detected as low misses. High misses are detected by photoelectric outputs in excess of the maximum signal permissable for a hit indication. The sight adapter superimposes a high or low miss indication light in the same relative position of the reticle sight pattern that a spotting round would normally appear.

Due to the relatively low velocity of the 106mm round, it is necessary to simulate lead on a moving target. This is accomplished by displacing the optical axis of the telescope through a fixed lead angle as the gunner maneuvers the traversing mechanism of the weapon.

Vehicle and Fortification Targets

Operation of the system with vehicle and fortifications targets is illustrated in Figure 8. Four flashlamp assemblies, attached to the vehicle, are so directed that flashes from only one lamp are visible in only one quadrant.

Each of these flashlamp assemblies contains 2 flash tubes, an IR filter, trigger coil circuits, a 12 volt rechargeable nickel cadmium battery and a 450 volt power supply. A supplementary pack is used in conjunction with the personnel target pack to operate the flash lamp assemblies mounted on the vehicle. An unfiltered xenon flashtube is also used for indication of small arms projectile hits. A pyrotechnic hit indicator is triggered only when the vehicle is hit by a large caliber weapon such as a 106mm recoilless rifle or a significant number of hits by small arms.

Conclusion

Although the system described is only in the experimental stage, the special techniques employed can be of considerable importance to the Army in their development of a future combat assessor system.

Credit

Appreciation is extended to Mr. Philip Knapp and others of Aircraft Armaments Inc. for their work in the design and development of this system.

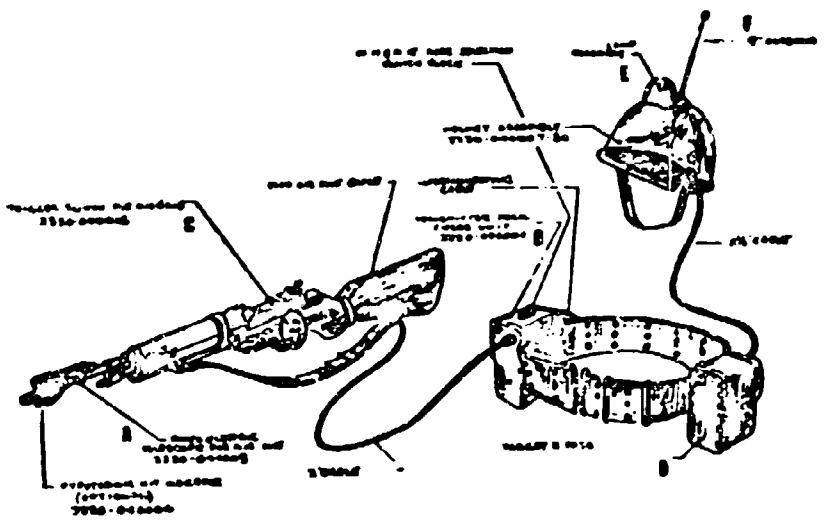


FIG. 1

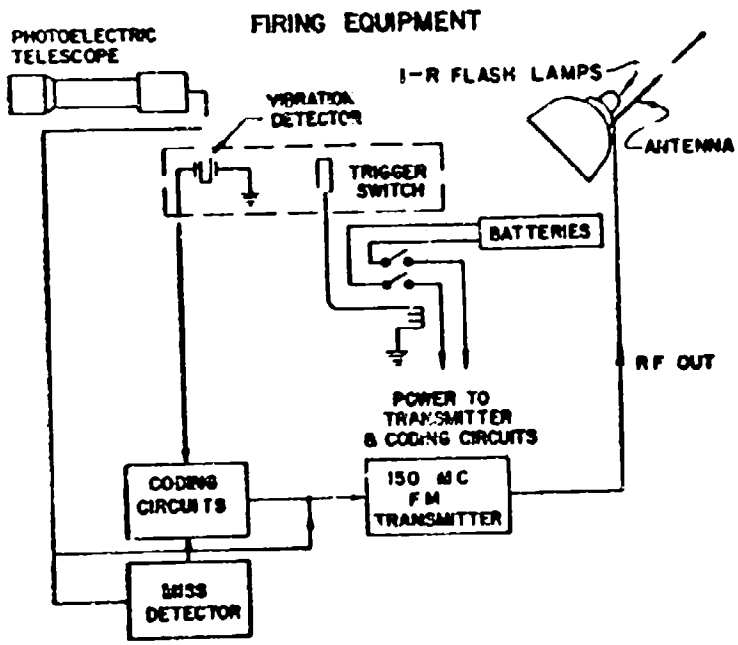

FIRING EQUIPMENT
PHOTOELECTRIC TELESCOPE
I-R FLASH LAMPS
VIBRATION DETECTOR
ANTENNA
TRIGGER SWITCH
BATTERIES
RF OUT
POWER TO TRANSMITTER & CODING CIRCUITS
CODING CIRCUITS
150 MC FM TRANSMITTER
MISS DETECTOR

FIG. 2A

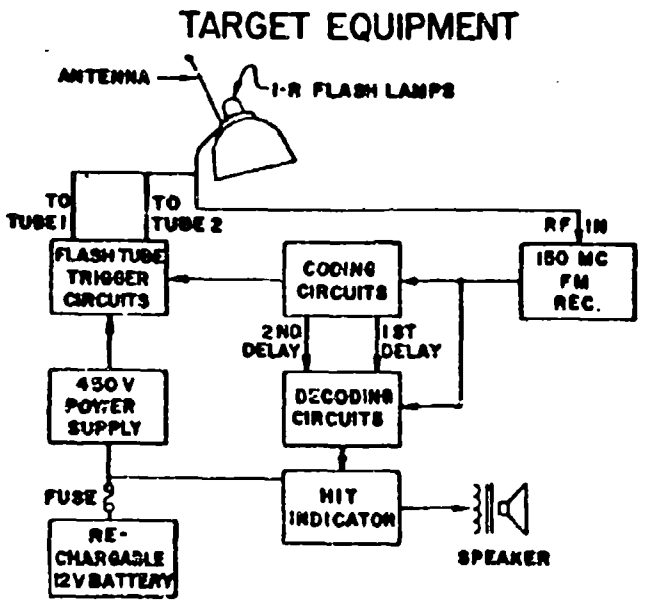

TARGET EQUIPMENT
ANTENNA
I-R FLASH LAMPS
TO TUBE 1
TO TUBE 2
RF IN
FLASH TUBE TRIGGER CIRCUITS
CODING CIRCUITS
150 MC FM REC.
2ND DELAY
1ST DELAY
450 V POWER SUPPLY
DECODING CIRCUITS
FUSE
HIT INDICATOR
RE-CHARGABLE 12V BATTERY
SPEAKER

FIG. 28

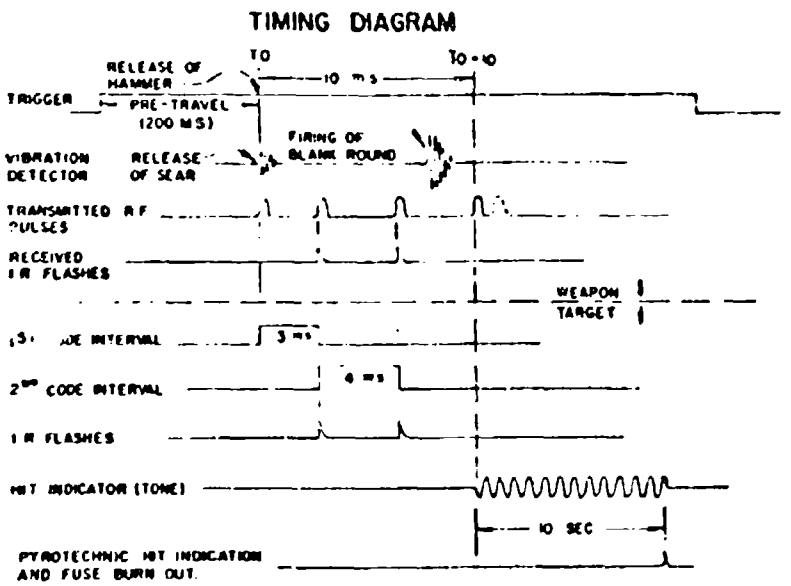

TIMING DIAGRAM
RELEASE OF HAMMER
T0
T0 + 10
10 m s
TRIGGER
PRE-TRAVEL (200 MS)
VIBRATION DETECTOR
RELEASE OF SEAR
FIRING OF BLANK ROUND
TRANSMITTED RF PULSES
RECEIVED I R FLASHES
WEAPON TARGET
1ST CODE INTERVAL
3 m s
2ND CODE INTERVAL
6 m s
I R FLASHES
HIT INDICATOR (TONE)
10 SEC
PYROTECHNIC HIT INDICATION AND FUSE BURN OUT.

FIG. 3

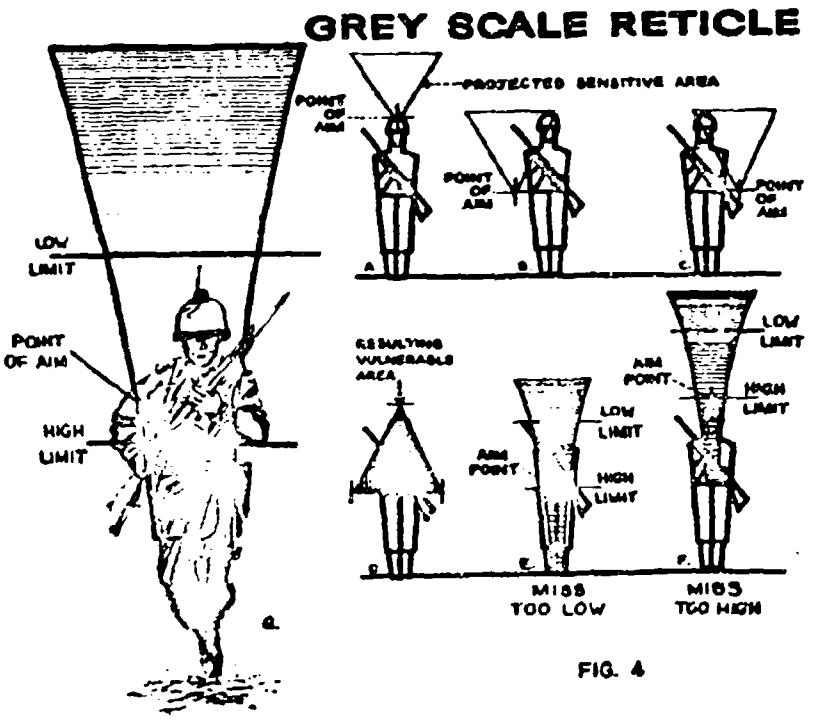
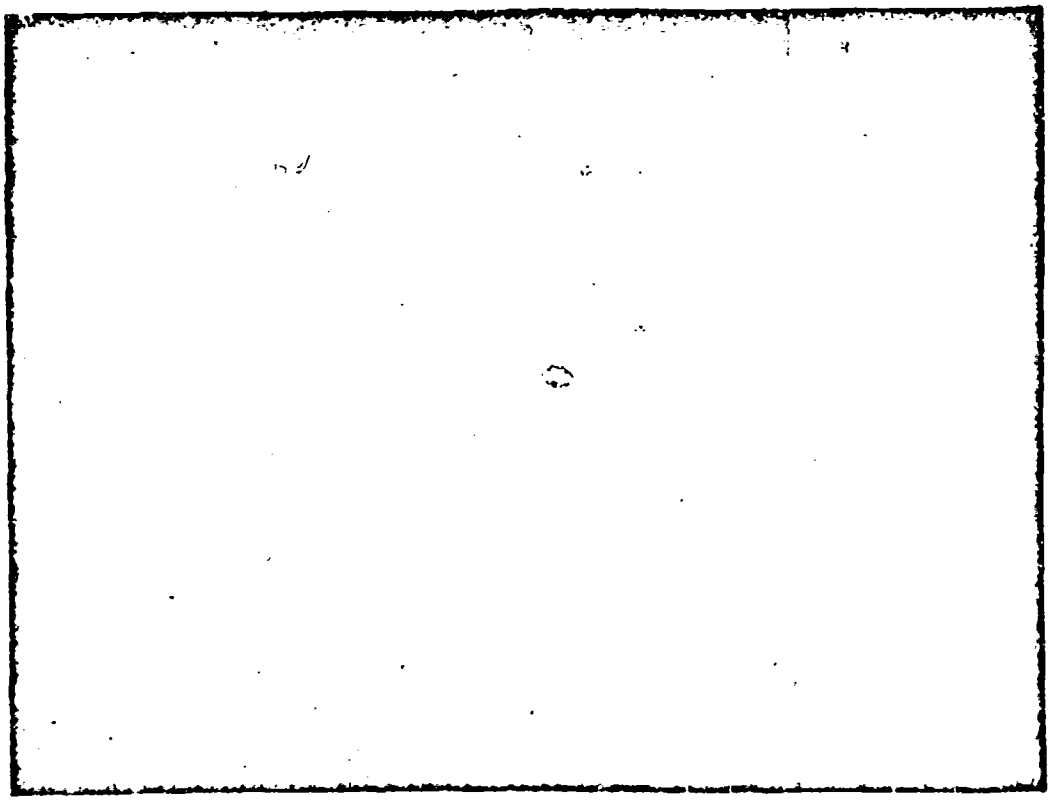

GREY SCALE RETICLE
PROJECTED SENSITIVE AREA
POINT OF AIM
LOW LIMIT
HIGH LIMIT
RESULTING VULNERABLE AREA
AIM POINT
MISS TOO LOW
MISS TOO HIGH

FIG. 4



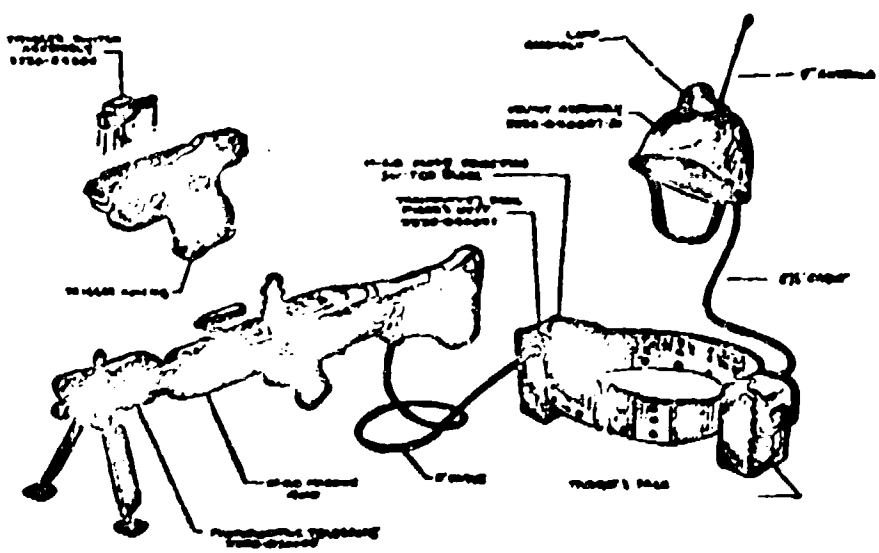


FIG. 7

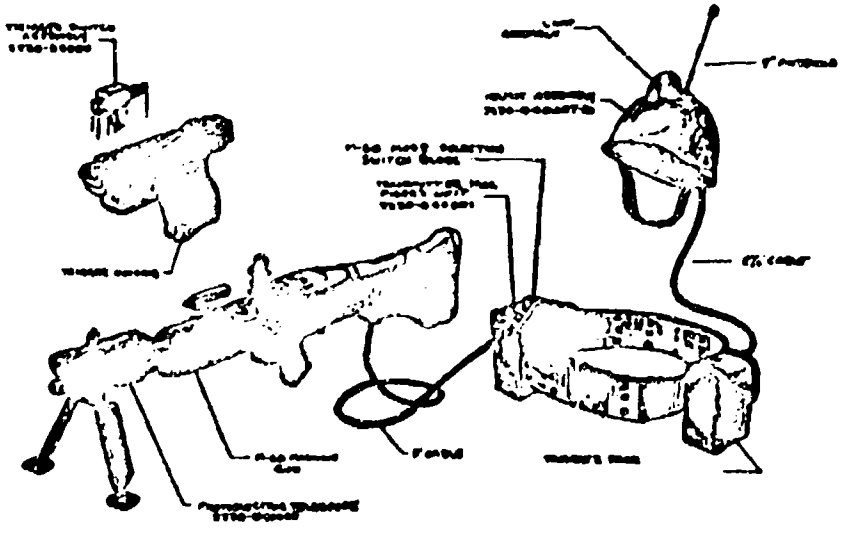


FIG. 7

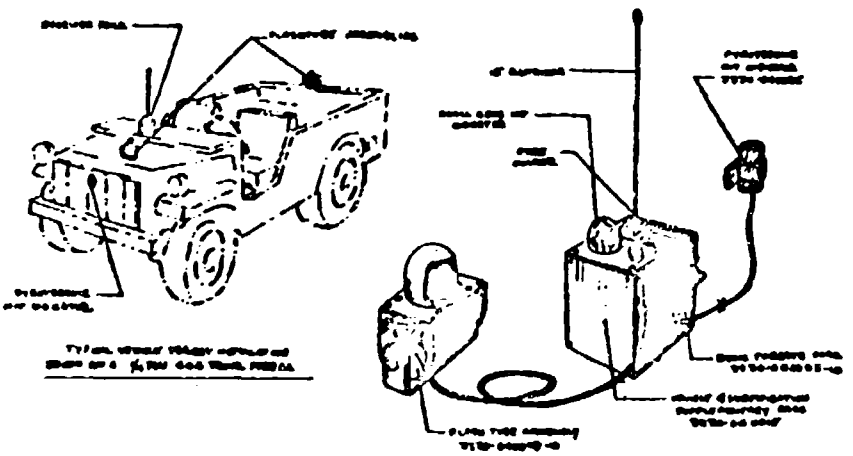


FIG. 8

VEHICLE AUX FORTIFICATION
TARGET EQUIPMENT

NOTES

A DELAY-LINE SYNTHESIZED FILTER BANK WITH ELECTRONICALLY ADJUSTABLE IMPULSIVE RESPONSE

H. J. Riekel
E. Brockner
Federal Scientific Corporation
615 West 131st Street
New York 27, New York

Introduction

In many radar and [illegible] tems useful information i[illegible] in the frequency of received [illegible] The signals of interest are u[illegible] b-scured by noise and banks o[illegible] ontiguous filters have been used for frequency identification and signal-to-noise ratio enhancement. However the use of such filters presents the following problems: 1) If good frequency resolution and signal enhancement are desired, a very large number of filters must be used; 2) The outputs of the filters are essentially simultaneous and to ascertain which filters respond to a certain input it is necessary to sample the outputs of all the filters in a relatively short time; 3) It may be desired to employ the transient characteristic of filters to achieve linear signal integration. If so the energy storage in the resonant filter elements limits the rapidity with which the signal processing can be repeated. These and other problems associated with the use of discrete filters are overcome by a delay-line synthesized filter bank which was developed by Federal Scientific Corporation for the U.S. Air Force under Contract AF 30(602)-1613.

The high-resolution spectrum analyzer for which the trademark "SIMORAMIC" has been adopted synthesizes the equivalent of thousands of bandpass filters located side by side in the frequency domain through the use of a single delay line in a well-controlled closed loop. The analyzer presents the results of its real time signal integration and spectrum analysis in panoramic form. It is not a swept filter, but it operates simultaneously over its entire coverage range. The frequency location, impulsive response, and gain of all the synthesized filters are determined by the same network elements within the closed loop, making for excellent reliability and stability. The outputs of the individual synthesized filters are presented sequentially for either human observation or computer consumption. The impulsive response of the individual synthesized filters can be readily changed electronically, permitting the system to adapt quickly to changing requirements of frequency resolution and integration efficiency.

Before discussing the operating principle and the characteristics of the SIMORAMIC Analyzer, it is pertinent to review the available methods of spectrum analysis and the analyzer characteristics required for the successful processing of transient signals in the presence of noise.

Spectrum Analysis of Time-Varying Signals

A time-varying signal has a frequency spectrum which changes as a function of time. Consider, for example, the problem of analyzing speech signals. During the pronounciation of a single word the sound spectrum changes quite rapidly and in speech analysis it is frequently desired to obtain many successive "short term" power spectra during the pronounciation of a single word. If a complete Fourier spectrum of a signal is to be obtained in real-time the analyzer must be able to observe all frequency components of the signal simultaneously.

There are many occasions when it is of the utmost importance to measure the spectrum of a signal in real-time. An example is afforded by a radar system which is called upon to distinguish between different types of targets by real-time determination of target scintillation characteristics. Similar problems exist in the field of sonar.

Heretofore, spectrum analysis was accomplished by instruments which, with few exceptions, fell into two broad classes: 1. the heterodyne-type analyzer and 2. instruments employing a multiplicity of contiguous filters.

The heterodyne-type analyzer utilizes a single narrow bandwidth filter which is resonant at a fixed frequency. The spectrum which is to be investigated is presented to the filter in a sequential manner by means of a frequency converter which heterodynes the signal spectrum to the fixed resonant frequency of the filter.

For the analysis of time-varying signals this device has serious disadvantages. The scanning of a spectrum under investigation consumes a relatively long period of time since the scanning speed must be slow enough to permit the resonant circuit to respond to any signal present. Further, there may be signals in the spectrum which persist only briefly. In that event it is likely that the signal will not be heterodyned to the filter resonant frequency at the correct time so that such signals will often remain undetected. For example, the time required to scan a 10,000-cps wide spectrum with a 5-cps filter is approximately 400 seconds. The difficulties mentioned above are completely eliminated in the SIMORAMIC Analyzer.

Some of the drawbacks of the heterodyne-type analyzer can be eliminated by use of time-compression techniques or at the expense of non-real-time operation.

The second type of analyzer employs a bank of contiguous filters. The analysis of a 10,000-cps spectrum with uniform 5-cps resolution requires approximately 2000 filters. Each of the filters constitutes a separate transmission channel and the gains and frequency characteristics of all 2000 filter channels must be carefully adjusted in order to achieve a constant gain versus frequency response in the analyzer. It is easy to see that gaps may develop in the frequency response of the filter bank if the selectivity characteristics of the individual filters do not overlap prpperly. If a visual indication of a time-varying signal spectrum is desired it is necessary to sample the outputs of all filters during one filter time constant. Thus for the example used previously, the outputs of all the 2000 filters must be examined every 200 milliseconds. (200 msec is the nominal time constant of a 5-cps wide filter.) The sampling problem is avoided in the SIMORAMIC Analyzer which automatically maps the spectrum of the input signal into time. Furthermore the SIMORAMIC Analyzer synthesizes a continuum of filters with identical selectivity characteristics. The shape of the selectivity characteristic can be controlled electronically.

Characteristics of the SIMORAMIC Analyzer

The basic block diagram of a SIMORAMIC Analyzer channel is shown in Fig. 1. It consists of an ultrasonic delay line and a heterodyner in a closely regulated unity gain feedback loop. The heterodyner may be thought of as a time-variable phase shifter, so that the network is time-variable. Amplifiers required for the implementation of the loop are not shown. Using a time-multiplex arrangement ten basic SIMORAMIC Analyzer channels can be built around a single delay line. Detailed analyses of the operation of the basic loop of Fig. 1 as a spectrum analyzer and coherent signal integrator have been presented previously.[1,2] A brief introduction to the operating principle of the device will be given before presenting a summary of its characteristics.

In the absence of the loop heterodyner, i.e., the time-varying phase shifter, the device of Fig. 1 reduces to the well-known comb filter. The characteristics of comb filters have been described by White and Ruvin.[3] The solid graph in Fig. 2 illustrates the frequency response of a comb filter. As indicated in the figure the device cannot distinguish between frequencies which are separated by $1/T$ cps. It is easy to show that the response frequencies of the basic delay line loop can be shifted by the addition of a static phase shifter in the loop. Thus by introducing a static phase lag of $\Delta\omega T$ radians in the loop, the "comb frequencies" become shifted by $\Delta\omega$ as indicated in the figure. However, even this

alteration of the basic delay line loop is not sufficient to convert the simple delay line integrator into a useful spectrum analyzer.

In order to convert the delay line integrator into a useful spectrum analyzer, the signal applied to it is first restrained to the frequency range

$$\frac{K}{T} \leq f < \frac{K+1}{T}$$

The temporal sampling theorem states that a signal whose spectrum is confined within a band 1/T cps wide, is uniquely determined by its complex values at intervals T. Thus on the basis of this theorem, it is not necessary to observe and process the bandlimited signal continuously but only its inspection once every T seconds is called for.

The heterodyner in the delay line loop of Fig. 1 "tunes" the filter through 1/T cps every T seconds. The period of this tuning is precisely equal to the loop delay and the resulting time-varying filter system may be represented as shown in Fig. 3. As shown in this figure, the signal whose spectrum has been confined to 1/T cps is applied sequentially to a bank of very narrow filters which cover the spectrum occupied by the signal. The rate of switching is such that the same filter is excited once every T seconds.

A summary of the characteristics of a basic SIMORAMIC Analyzer channel will be presented next. To make the discussion numerical, it will be assumed that a storage loop delay, T, of 1000 microseconds is used and that the time of signal processing is 200 milliseconds. Under these conditions, one cycle of analyzer operation proceeds as shown in Fig. 4. The salient characteristics of the analyzer may be stated as follows:

1. In order to assure unambiguous operation the signal spectrum applied to the basic storage loop is limited to a band of 1/T cps, or in the case at hand, 1000 cps. Time-multiplexing of one basic storage loop is employed to increase the frequency coverage to 10 kc. However, in the discussion which follows, only the "basic" channel characteristics will be referred to.

2. The particular band of frequencies covered by the SIMORAMIC Analyzer may be positioned at any desired frequency.

3. The signal is applied to the analyzer for a processing period, P, of 200 milliseconds.

4. The output is observed during the last millisecond of signal application. During the output observation period, a panoramic display of the analysis of 1000 cps of signal spectrum is obtained.

5. The output observation period is followed by a clearing period of approximately 2 milliseconds immediately after which the cycle is repeated.

6. During the observation period, the panoramic display resulting from three simultaneous sinusoidal input signals may appear as in Fig. 5. Each one of the pulse-like waveforms is called a response function; it is in fact, the selectivity characteristic of the instrument.

7. The peak in the response function, as observed during the 1000-microsecond observation period, occurs at an instant which is linearly related to the frequency of the input signal.

8. The amplitude of the response peak is directly proportional to the amplitude of the applied signal.

9. The uncertainty in identifying the frequency of a sinusoidal input component is inversely proportional to the period of signal processing, P. As used here, uncertainty approximately equals 1/200 milliseconds or about 5 cps.

10. The shape of the basic selectivity characteristic (response function) is illustrated in the oscillogram of Fig. 6a. It is of the form $\frac{\sin x}{x}$. One of the unique features of the SIMORAMIC Analyzer is the ease with which the selectivity characteristic can be altered to suit a particular application. The shape of the output, illustrated in the

figure, was obtained by applying to the device a constant amplitude sine wave. If however, the input signal is weighted by amplitude modulation during the processing time, a drastic change occurs in the shape of the response function. This situation is best illustrated by the oscillogram of Fig. 6b which shows the instrument's output in response to a cosine weighted noise-free sinusoid. It should be noted that the lobes in the response pattern are eliminated at the cost of a slight broadening of the response peak. The analysis of the operation of the device with amplitude-weighted inputs parallels the theory of amplitude-weighted linear antenna arrays. An alternate point of view is provided by attributing to the filters synthesized by the delay line loop an "effective" impulsive response which can be varied by time-weighting.

The Synthesis of an Arbitrary Bank of Filters

The SIMORAMIC Analyzer can synthesize the selectivity characteristics of an arbitrary bank of filters as shown in Fig. 3. The sampling rate at the input of the equivalent filter bank is very high compared to the nominal time constant of the individual filters and under this condition the presence of the sampling switch may be neglected in the analysis whose outline follows.

The impulsive response of an arbitrary conventional bandpass filter can be expressed in the form

$$h(t) = c(t)\cos\left[2\pi f_o t + \phi(t)\right] \qquad (1)$$

where $c(t)$ = envelope of impulsive response

$\phi(t)$ = phase modulation function

and f_o = filter's center frequency and carrier frequency of impulsive response

For narrow-band filters, $c(t)$ and $\phi(t)$ are slowly varying functions compared to the carrier. Homogeneous filters are characterized by impulsive responses having identical $c(t)$ and $\phi(t)$.

Let the signal applied to the filter be expressed in the form

$$e_i(t, \Delta f) = \sin\left[2\pi(f_o + \Delta f)t + \Delta\right] \qquad (2)$$

where Δf = deviation of signal frequency from the center frequency of the filter

Δ = signal phase angle

Then the response of the filter at time $t = P$ is given by

$$e_{out}(P, \Delta f) = \int_0^P h(t)\, e_i(P-t, \Delta f)\, dt \qquad (3)$$

The selectivity characteristic of the filter is defined here as the envelope of $e_{out}(P, \Delta f)$ as a function of Δf.

It can be shown[4] that for $\Delta f \ll f_o$ the filter selectivity characteristic is given by

$$H(\Delta f) \cong \left| \int_0^P \tfrac{1}{2} c(t) \sin\left[2\pi \Delta f t - \phi(t)\right] dt + j \int_0^P \tfrac{1}{2} c(t) \cos\left[2\pi \Delta f t - \phi(t)\right] dt \right| \qquad (4)$$

$$\text{for } \Delta f \ll f_o$$

In order to make the selectivity characteristic of a delay-line synthesized filter identical to that of a desired arbitrary conventional filter the basic SIMORAMIC Analyzer must be modified as shown in Fig. 7. At the output of the delay line synthesized filter bank frequency is mapped into time. Thus the selectivity characteristic of the device is obtained by determination of its time response to a single sinusoid.

Apply the signal of Eq. 2 to the SIMORAMIC Analyzer of Fig. 7. The output of the system may be found by employing

the approach of Reference 1 and approximating the summations by integrals. It is shown on page 104 of Reference 4 that the selectivity characteristic of the modified SIMORAMIC Analyzer may be expressed in the form

$$H(\Delta f)_{mod} = \left| \frac{1}{T} \int_0^P c(x) \cos\left[2\pi \Delta f x - \phi(x)\right] dx + j \frac{1}{T} \int_0^P c(x) \sin\left[2\pi \Delta f x - \phi(x)\right] dx \right| \quad (5)$$

Equation 5 is, except for a coefficient, identical to Eq. 4 which expressed the selectivity characteristic of a conventional filter in terms of its impulsive response.

The input signal to the basic SIMORAMIC Analyzer is amplitude-weighted in accordance with the mirror image of c(t) and phase-modulated in accordance with the mirror image of $\phi(t)$. By means of this technique, it is possible to synthesize the frequency selectivity characteristics of an arbitrary bank of filters. The synthesis procedure leads to excellent results if the functions c(t) and $\phi(t)$ vary little during T, the storage loop delay time. If c(t) is made unity and $\phi(t)$ equals zero, the modified system reduces to the basic SIMORAMIC Analyzer which synthesizes a bank of ideal coherent integrators. It is important to note that the device does not synthesize the desired bank of filters at all times. Only during the brief observation interval at the end of each integration period P (see Fig. 4) is the SIMORAMIC Analyzer output equivalent to the output of a conventional filter bank. However, in using a bank of filters for the purpose of Fourier analysis and signal detection this is the only output required.

Because the loop delay T is much smaller than the signal processing time P (typically $T = \frac{1}{100} P$), the selectivity characteristics of all filters synthesized by the time-varying delay line system are of identical shape. Further, the gains of all filters are the same since the filters are synthesized by the same network elements within the closed loop.

Figure 8 shows a few salient filter bank characteristics which can be realized with the SIMORAMIC Analyzer. The column labeled "60 db point for P = 1 sec" indicates the displacement from center frequency at which the selectivity characteristic of a nominal 1-cps wide filter is down 60 db. With rectangular weighting the delay line system synthesizes a bank of ideal coherent integrators and attains theoretically optimum signal-to-noise enhancement of sinusoidal signals. Its resolution bandwidth is also optimum under these conditions. However, the selectivity characteristic of the coherent integrator has a lobe structure which is undesirable if small signals must be detected in the presence of large ones. The right-hand column in Fig. 8 indicates the magnitude of the largest lobes of the various selectivity characteristics. It is evident that a selectivity characteristic with rapid attenuation off center frequency can be attained only at the expense of some loss in signal enhancement and resolution.

Applications

Figure 9 shows an interesting application of a SIMORAMIC Analyzer. In this figure are shown the typical responses of an analyzer which was completed recently to each of two ostensibly identical 40-mc oscillators. Each frame shows the signal spectrum on a frequency scale of 85 cps per box. The separate frames within each group were taken a short time interval apart. The oscillator spectrum shown on the left exhibits a broader central response than the one on the right. The responses shown on the right are richer in low-frequency sideband components. After some practice, it is possible to identify the oscillators simply by examination of successive outputs of the analyzer. Thus, the SIMORAMIC analyzer can be used to identify signal sources by analyzing and presenting their characteristics or "signatures."

The system used in the above-mentioned experiment is capable of analyzing a 10-kc wide spectrum with 6-cps resolution. Such a system incorporates about 300 tubes.

Other system parameters are possible. For example, a system has been designed which analyzes a spectrum ranging from dc to 200 cps with 1-cps resolution. In general, the linear processing range of the SIMORAMIC Analyzers is between 40 db and 60 db, and the linearity of the frequency display axis (frequency-time conversion) is approximately 1 part in 100,000. The absolute frequency accuracy typically is 1/5 to 1/10 of the nominal frequency resolution.

The SIMORAMIC Analyzer is useful in the analysis of vibration, of sound, in the determination of Doppler frequencies and, in general, wherever there is need for repeated linear integration and fine resolution spectrum analysis.

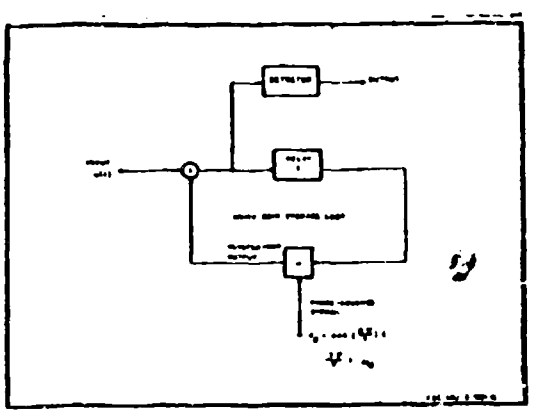
References

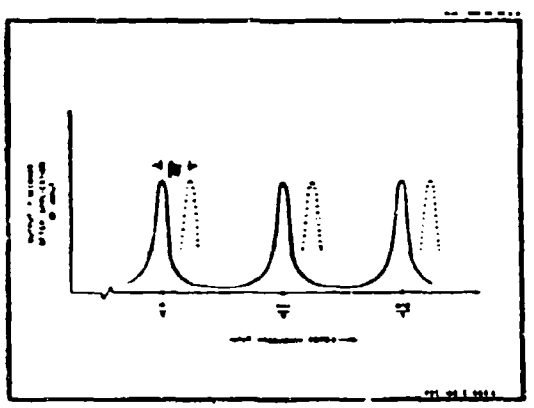
1. Henry J. Bickel, "Spectrum Analysis with Delay-Line Filters," 1959 IRE WESCON Convention Record (Part 8).

2. Jack Capon, "On the Properties of an Active Time-Variable Network: The Coherent Memory Filter," Proceedings of the Symposium on Active Networks and Feedback Systems, Polytechnic Institute of Brooklyn, 1960.

3. Warren D. White and A.R. Ruvin, "Recent Advances in the Synthesis of Comb Filters," 1957 IRE National Convention Record (Part 2).

4. E. Brookner, "The Synthesis of Arbitrary Banks of Filters," ITM No. 49, dated June 1958 (unpublished memorandum). Federal Scientific Corporation, New York 27, N. Y., or Progress Report FSC-P-5/101, Accuracy Improvement Techniques for Long Range Radar, (Title Unclassified), Federal Scientific Corporation, New York 27, N. Y. 1 August 1958, ASTIA No. AD-148908. SECRET.





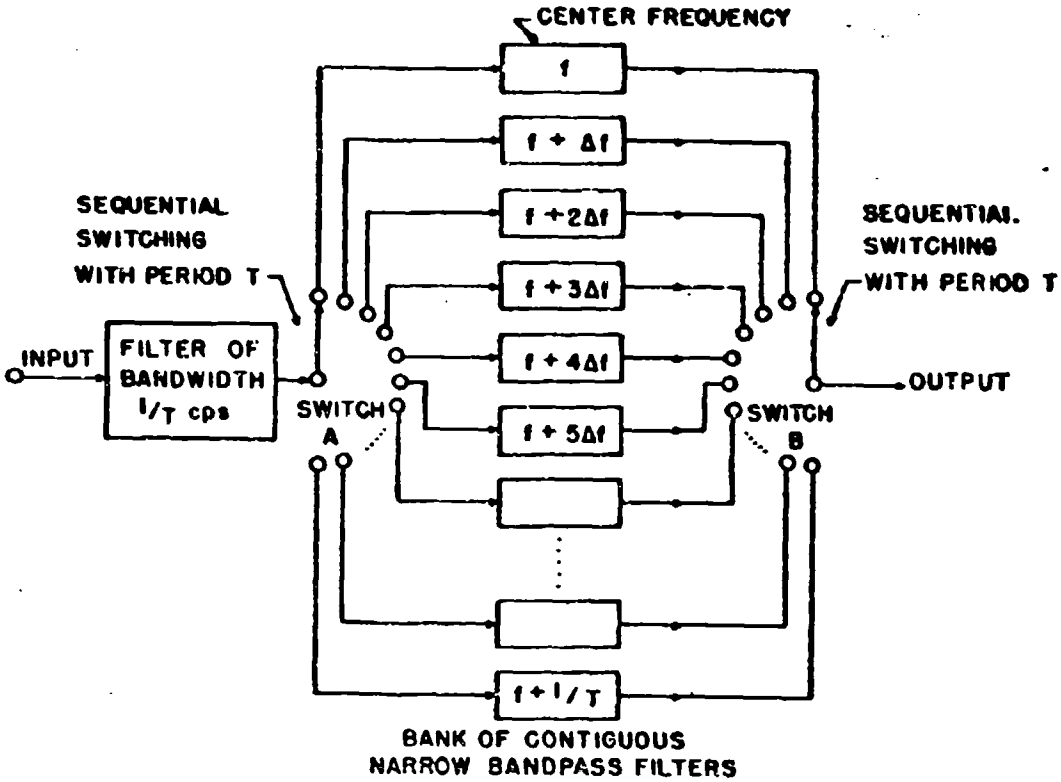
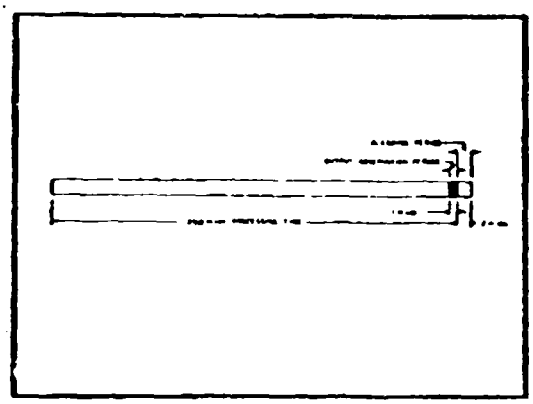


FIG. 3 SEQUENTIALLY EXCITED AND SAMPLED BANK OF FILTERS



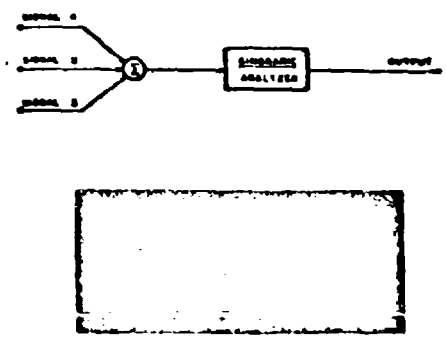
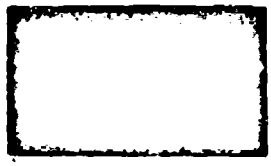
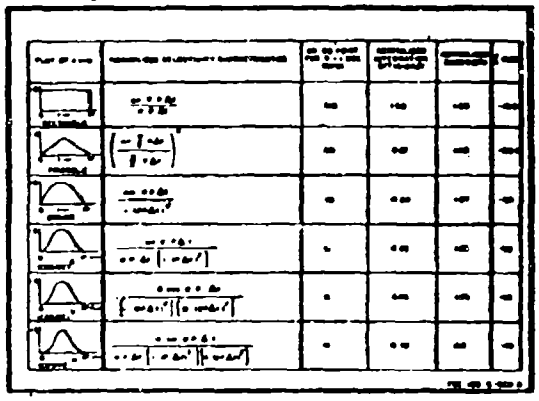
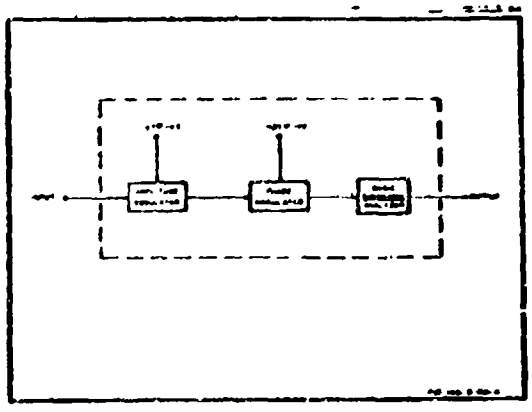


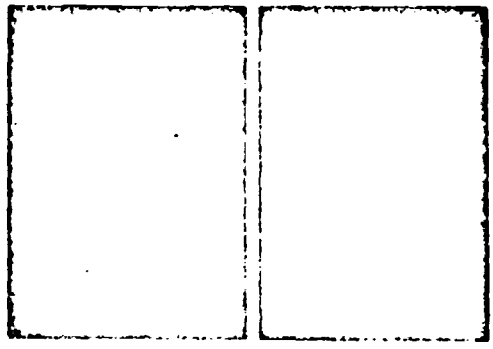
FIG 5 SPECTRUM ANALYZER DISPLAY FOR THREE SIMULTANEOUS INPUT SIGNALS











SPECTRUM OF INSTRUMENT I

SPECTRUM OF INSTRUMENT II

FIG. 9 OSCILLOGRAMS SHOWING TIME-VARYING SPECTRA OF TWO OSTENSIBLY IDENTICAL INSTRUMENTS

INSTRUMENTATION USED FOR IONOSPHERE ELECTRON DENSITY MEASUREMENTS

By: Mr. William J. Cruickshank, Ballistic Measurements Laboratory,
Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

ABSTRACT

The technique employed by the Ballistic Research Laboratories for the measurement of ionospheric electron densities through study of differential propagation effects on two phase coherent frequencies broadcast from high altitudes is discussed. A brief description of the stable airborne transmitters and of the rocket vehicle designated "Strongarm" which was used in an initial experiment to carry the airborne transmitters to an altitude of 1,000 miles is presented. The very narrow band receivers and recording equipment needed to provide data reception and information separation are discussed. A summary is given of the results of the two "Strongarm" rockets that have been fired, and the schedule of future experiments is outlined.

INTRODUCTION

Since 1946 the Ballistic Research Laboratories have been instrumenting upper atmosphere research rockets and guided missiles with a highly precise radio doppler tracking system known as DOVAP[1,2]. Beginning with the first of the Soviet Sputniks, these Laboratories have also been tracking the majority of artificial earth satellites and space probes through the use of radio doppler measurements. One source of error in obtaining position data from these systems is the refraction of radio waves as they are transmitted through the ionosphere. This error is related to the equivalent electron density and electron content of the ionosphere. Since very little is known about the electron density profile above the F2 region, a joint effort by the Ballistic Research Laboratories and the University of Michigan was undertaken to make measurements of this profile. The technique used for obtaining this profile employed measurements of the differential propagation effects on two harmonically related radio signals as they are transmitted through the ionosphere. The two signal frequencies used were 37 mc/s and 148 mc/s with the latter used as a reference since it is relatively unaffected by the ionosphere. A sounding rocket launched in a near vertical trajectory was used to carry the radio frequency transmitters aloft, and the measuring instrumentation was located on the ground below the rocket.

Measurements of nose cone temperature and vehicle performance were also objectives of the experiment.

The experiment was conducted at the National Aeronautics and Space Administration's facility on Wallops Island, Virginia, and secondary ground receiving equipments were located at the Ballistic Research Laboratories' tracking station on Spesutia Island at the Aberdeen Proving Ground.

INSTRUMENTATION

Rocket Vehicle

The rocket vehicle, designated "Strongarm", used for the experiment, was a five stage combination of solid propellant rockets and rocket boosters. This configuration, which is 56.5 feet long and weighs 7120 pounds, is composed of the following: an "Honest John" first stage, two "Nike-Ajax" boosters for the second and third stages, a "Yardbird" for the fourth stage, and a scaled "Sergeant" for the fifth stage. The above combination with a 20 pound payload was expected to attain a peak velocity of approximately 17,000 feet per second, a burnout altitude of 200,000 feet, and a peak altitude in excess of 1,000 miles. A photograph of the rocket combination on the launcher is shown in Figure 1. Development of the rocket combination was accomplished by the University of Michigan under contract to the Ballistic Research Laboratories.

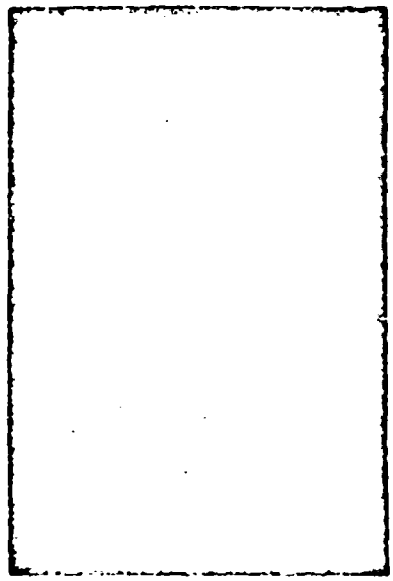


Figure 1. Strongarm Rocket Vehicle

Airborne Instrumentation

The payload package that was used for the first "Strongarm" flight consisted of two transmitters and their associated antennas, a telemetry sub-carrier oscillator, temperature sensor, and an ablative nosecone housing. A pictorial sketch of this configuration is shown in Figure 2. This equipment was also developed by the University of Michigan for the Ballistic Research Laboratories [3].

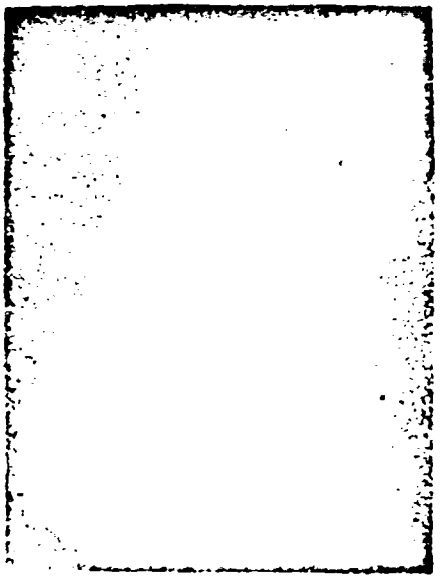


Figure 2. Airborne Payload Instrumentation

The heart of the airborne instrumentation was an extremely stable crystal controlled transistor oscillator used to drive both transmitters. Oscillator stability was maintained by precise regulation of the oscillator temperature to within $\pm 0.1^{\circ}C$. This temperature regulation was accomplished by using the heat of fusion method. The variance in the oscillator output frequency was held to less than one part in 10^{6} throughout the flight. The transmitters were small transistorized units having power outputs of 100 milliwatts and 20 milliwatts at 37 mc/s and 148 mc/s respectively. Figure 3 shows the two transmitters mounted on their base plate and the oscillator in its aluminum case.

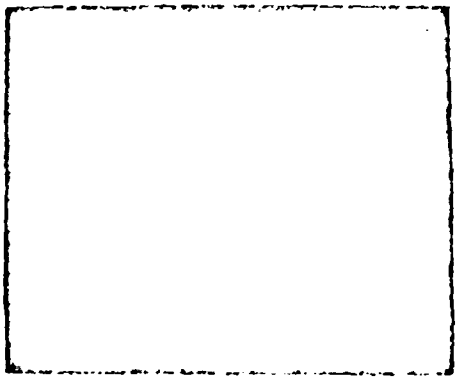


Figure 3. Airborne Oscillator and Transmitters

The antennas for these transmitters were two trapezoidal loops located in mutually perpendicular planes so that the radiation pattern of one loop antenna was essentially unaffected by the other. The antenna configuration, which was based on a missile loop antenna designed at the Ballistic Research Laboratories, is shown in Figure 4. Each antenna had an omnidirectional pattern in the plane of the loop and a figure eight pattern in the plane perpendicular to the loop. The efficiency of the 37 mc/s antenna was 16 db below an isotropic radiator, and that of the 148 mc/s antenna was 1 db below isotropic.

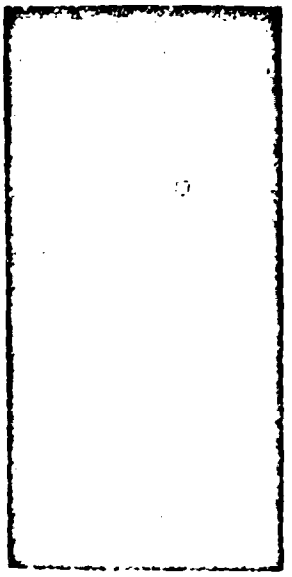


Figure 4. Airborne Antennas

The ablative nosecone, used to house the airborne instrumentation, consisted of a fiberglass shell with an outer teflon coating. This construction permitted radiation from the internal antennas, and kept the equipment from overheating. A temperature measurement at the interface of the fiberglass and teflon was desired; therefore, a thermistor temperature sensor and telemetry subcarrier oscillator were used to make this measurement. Input for the sub-carrier oscillator was provided by the sensor placed at the fiberglass-teflon interface. The pulsed output of this oscillator keyed the 148 mc/s transmitter at a rate that varied between 2 and 60 pps.

Ground Instrumentation

The primary ground reception and recording equipment was located at Wallops Island in a mobile ground station. This station was composed of a large complex of precision equipment in an 8 x 26 foot semi-trailer, and an array of six antennas in an adjacent field. A pictorial layout of the equipment in this trailer can be seen

in Figure 5. An auxiliary ground station was located 110 miles away at Spesutie Island. This station presently houses the Laboratories' satellite tracking facility.

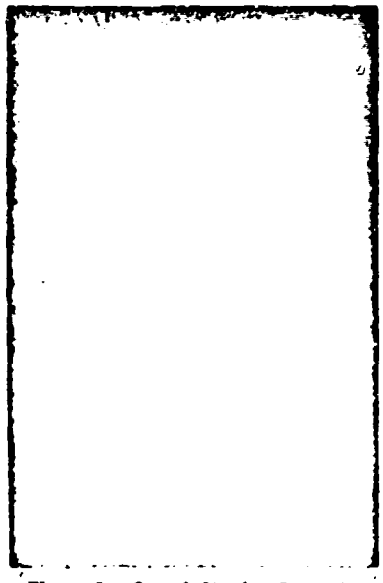


Figure 5. Ground [illegible] and Receiving Equipment

System Explanation. A simplified block diagram of the ground instrumentation is shown in Figure 6. Each radio frequency (rf) signal transmitted from the moving rocket is received by a pair of helical antennas having circular polarity of opposite sense. This antenna configuration allows frequency errors caused by rocket spin and Faraday effect to add to the biased doppler frequency in one of the dual channels, and to subtract from it in the other dual channel. After the two harmonically related signals are received and filtered, they are applied to differential mixers. The output of these mixers contains twice the spin frequency and Faraday effect. This error output from the two harmonically related signals is again mixed to give the Faraday differential frequency. This frequency is used to compute the total electron content of the atmosphere between the rocket borne transmitting and ground receiving antennas. The signals from the tracking filters are also applied to adder networks in a similar combination to that above; thereby, eliminating spin and Faraday rotation error at each of the two rf frequencies. The 37 mc/s adder output, which is a 7 kc/s bias plus twice the doppler frequency, is multiplied by two. The 148 mc/s adder output, which is a 28 kc/s bias plus twice the doppler frequency, is divided by two. After this multiplication and division is performed, the resultant frequencies are compared in a differential mixer. The output of this differential mixer is the doppler difference or dispersive doppler that is used to compute electron density.

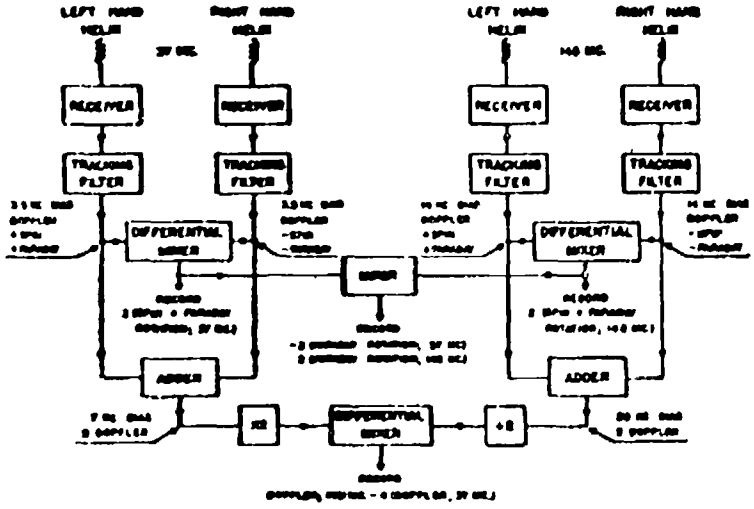


FIGURE 6 IONOSPHERIC PROBE - GROUND STATION BLOCK DIAGRAM

Antennas. The pairs of helical antennas used on the ground have desirable features such as circular polarization and relatively high gain[4]. The characteristics of the 37 mc/s units are as follows: 2.5 turns, 8.5 feet in diameter, 14.5 feet in length, 70° beamwidth, and a gain of 6 db over an isotropic radiator for linearly polarized signals. The 148 mc/s antennas have 8 turns, a diameter of two feet, a length of 12 feet, a beamwidth of 40°, and a gain of 11 db over an isotropic radiator. In the experiment here being discussed, each pair of antennas was placed four wavelengths apart at its particular frequency. Both pairs of antennas were placed in a single line perpendicular to the rocket firing azimuth. This placement cancelled parallax effects that would be encountered in frequency reception.

In addition to the four main helical antennas there was a helical antenna for reception at 148 mc/s and a half-wave dipole antenna for reception at 37 mc/s. The particular fuction of these two additional antennas will be explained in a later section.

Receivers and Pre-Amplifiers. The ground receivers, used for the initial experiment, were designed and built at the Ballistic Research Laboratories for use specifically for dispersive doppler measurements. Referring to the existing ground station, there are four triple conversion superhetrodyne receivers; all driven by common local oscillators through appropriate multipliers and isolation networks. A block diagram of the receivers is shown in Figure 7. This triple conversion superhetrodyne system is used for increased stability, high gain, and better signal to noise ratio at the output. The first two local oscillators are crystal controlled and have a stability of one part in 10^7 after 30 minutes warm up time. The third local oscillator is a highly stable variable unit used to precisely set the bias frequency. This variable oscillator is necessary because the various mixers and adders used for data handling depend upon a bias of 3.5 kc/s ± 1% for the low frequency doppler and 14 kc/s ± 4% for the high frequency doppler. The stability of the local oscillators is essential for keeping the bias frequencies within tolerance; however, the data outputs (Figure 6) are not dependent upon local oscillator or rf frequency stability since any such shifts are cancelled out. The bandwidths of the receivers are 5 kc/s and 20 kc/s for 37 mc/s and 148 mc/s, respectively.

In the field experiment, pre-amplifiers were used at each antenna. This combination of pre-amplifiers and receivers resulted in an approximate system input sensitivity of 0.1 microvolts for a receiver output signal to noise ratio of unity.

Tracking Filters. Audio electronic tracking filters (6) were used on the output of each receiver channel to obtain greater improvement in the signal to noise ratio. These devices are very narrow bandpass filters in which the output frequency follows or tracks the input frequency. This is accomplished automatically by the use of a high gain phase locked servo-controlled loop. Large signal to noise improvement in the output, as compared to the input, is obtained. The bandwidth used is dependent upon the dynamic change of the input frequency. Bandwidths of 50 cps and 25 cps were used on the high and low frequency channels respectively during the burning phases of the Strongarm rocket; 5 cps and 2.5 cps were

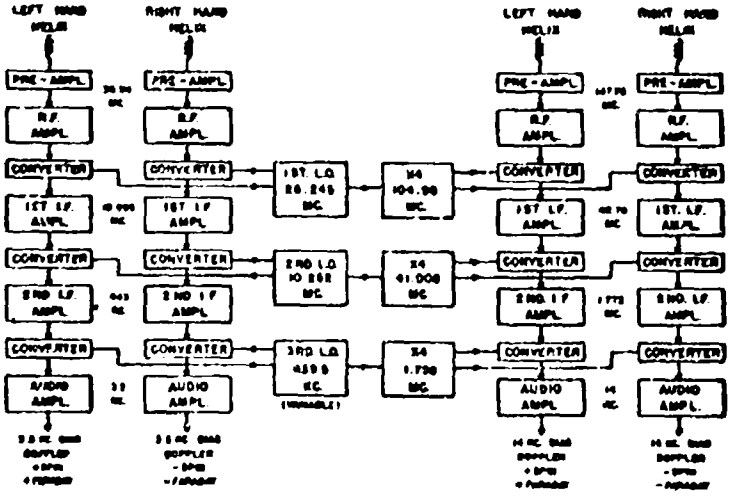


FIGURE 7 IONOSPHERE PROBE - BLOCK DIAGRAM OF RECEIVERS

used thereafter. The filter enabled tracking of signals to approximately 30 db below the receiver output noise, thereby, giving an overall system sensitivity of 3 nanovolts (3×10^{-9} volts).

Another feature of the filters is a constant amplitude output which is necessary for the adding and differential data networks.

Data Separation and Recording. The data circuits in the instrumentation system are composed of adders, differencing networks, mixers, multipliers, and dividers. The outputs of the above circuits, which are the data of interest, are: Spin and Faraday rotation at 37 mc/s, Spin and Faraday rotation at 148 mc/s, Faraday Difference and Doppler Difference. These data are recorded on an 8 channel chart recorder along with the Automatic Gain Control (AGC) output from each tracking filter. The tracking filter AGC output indicates the received signal level, even though operating below receiver noise level. It also gives definite indication of non-lock of the tracking filters. This is necessary since one non-locked tracking filter causes three out of the four chart recorded data channels to be useless. Magnetic tape recording is used for backup. In the experiment described, recordings were made of the receiver outputs, tracking filter outputs, and the two biased doppler frequencies from the adder circuits. A master frequency standard and timing generator, synchronized to the NBS station WWV, was used as the source for all frequency, period, and timing measurements. Local standard time in hours, minutes and seconds was recorded on all records.

Telemetry and Miscellaneous Instrumentation. The telemetry data from the "Strongarm" rocket were received and recorded at the Wallops Island and Spesutia Island stations. Since these data were encoded in a pulse amplitude modulated signal, a conventional AM receiver was used for recovering the data. A 148 mc/s helical antenna and pre-amplifier, identical to those previously described, were used with the AM receiver. Increased signal to noise ratio at the receiver output was achieved by restricting the output bandwidth with a low pass filter.

In order to resolve any possible ambiguity in determination of rocket spin rates, it was thought desirable to determine rocket spin rates independently of the system previously described. The rocket antennas radiate linearly polarized signals toward the ground stations and a dipole on the ground receives linearly polarized signals. The received rf amplitude variations resulting from this polarization effect is an excellent indication of rocket spin rate. Therefore, a system which consisted of a 37 mc/s dipole antenna, pre-amplifier and receiver was set up to obtain these data.

Each of the adder networks in the data circuits contain output signal frequencies which are twice the biased doppler frequency. These output frequencies are divided by two resulting in a low channel biased doppler frequency (3.5 kc/s ± 625 cps) and a high channel biased doppler frequency (14 kc/s ± 2500 cps). A "quick look" plot of doppler frequency versus time can be taken by integrating the low channel biased doppler frequency and recording this analog frequency output on a slow speed chart recorder. When the bias is subtracted, a true plot of doppler frequency versus time is presented. A printed tabulation of the period for the high channel biased doppler frequency was made at one second intervals prior to and during the rocket flight.

EXPERIMENTAL RESULTS

Rocket Vehicle

The first "Strongarm" rocket was successfully flown on 10 November 1959, at 0700 EST. All stages functioned normally and a maximum velocity of 17,100 feet/second was reached at the fifth stage burnout altitude of 215,000 feet. Peak altitude was in excess of 1,050 miles. This is the highest altitude attained to date by a firing from Wallops Island. Impact occurred at about 27 minutes after launching. A 20 minute flight was anticipated. This was the first firing of this combination of rockets and rocket boosters. All of the airborne instrumentation functioned well and reception at the ground receiving stations was excellent.

The second "Strongarm" rocket was launched on 18 November 1959, but no data were obtained. The last three stages of the rocket combination failed to ignite, and the peak altitude reached by the payload was only 20 miles as determined by "skin" tracking radar.

Electron Density Profile

Figure 8 shows a portion of the chart record from "Strongarm I", containing the four low frequency data outputs and timing. These data were used to compute the electron density and total electron content in the ionosphere. A profile of the electron density versus altitude, computed from preliminary data, is shown in Figure 9. This profile will be extended to 1600 km (1,000 miles) when all of the data have been reduced.



Figure 8

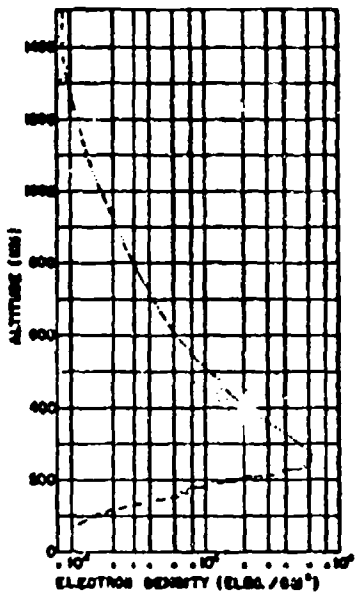


Figure 9. Electron Density versus Altitude

Nosecone Temperature

Results of the nosecone interface temperature measurements can be seen in Figure 10. Unfortunately an open thermistor circuit was present for the first 76 seconds of flight but after that the circuit recovered and valid data were received. A maximum temperature of 318°F was reached at about 100 seconds.

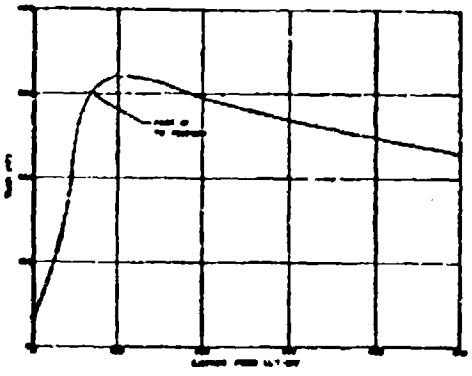


Figure 10. Nose Cone Temperature

Ionosphere Attenuation and Voltage Breakdown

Severe attenuation of both frequencies from the rocket transmitters was expected from shock-wave absorption associated with high mach numbers at low altitudes. The effect of this attenuation was not noted at either the Wallops Island station or at the Spocutie Island station. A slight attenuation of the 37 mc/s signal did occur for about 10 seconds at fifth stage burnout, but an increase in the 148 mc/s signal occurred for the same period of time. This phenomenon was attributed to rf breakdown of the air within the nosecone. The rf voltage at the upper gap of the 37 mc/s antenna had 52.5 volts across it, and this was adequate to initiate a glow discharge(7).

SUMMARY

The Ballistic Research Laboratories' "Strongarm" rocket was one of the first vehicles to carry instrumentation for making measurements of the ionosphere above the F2 region, and it was the first to make these measurements up to 1600 kilometers. Future experiments, which will use the same instrumentation are planned. Profiles of the ionosphere at noon and midnight are desired.

LIST OF REFERENCES

1. Zanoanato, M. W., "Ballistic Instrumentation for the IGY Rocket Project at Fort Churchill". Aberdeen Proving Ground, BRL Memorandum No. 1094, September 1957.

2. Zanoanato, M. W., "Ballistic Instrumentation and Summary of Instrumentation Results for the IGY Rocket Project at Fort Churchill". Aberdeen Proving Ground, BRL Report 1091, January 1960.

3. Orr, L. W., Cath, P. G., and Darnall, R. R., "A Two-Frequency Beacon for High-Altitude Ionosphere Rocket Research". Ann Arbor, RRI R-2616:3-3-F, December 1959.

4. Richer, K. A., "DOVAP Antennas for Pre-IGY". Aberdeen Proving Ground, BRL Memorandum 1067, July 1957.

5. Patterson, K. H., "A Pre-Amplifier Design for Satellite Receivers". Aberdeen Proving Ground, BRL Memorandum 1154, July 1958.

6. Richard, V. W., "DOPLOC Tracking Filter", Aberdeen Proving Ground, BRL Memorandum 1173, October 1958.

7. Paahs, H. A., "VHF Breakdown of Air at Low Pressures". Aberdeen Proving Ground, BRL Report 944, August 1955.

EFFECT OF ANTENNA PHASE PATTERN ON DOPPLER SYSTEM OPERATION*

By: H. S. Rothman and W. E. Scharfman, Stanford Research Institute

Introduction

To provide large improvements of the signal-to-noise ratio of noisy Doppler signals, various Doppler tracking systems[1,2,3] are employing phase-lock techniques. These techniques obtain maximum signal-to-noise by allowing an extreme reduction of bandwidth. The minimum allowable bandwidth is dependent on the maximum received rate of change of Doppler frequency.[4]

In this paper an analysis is carried out to determine the effect of the antenna phase pattern on the instantaneous frequency shift and rate of change of frequency shift of a received Doppler signal as a satellite passes through the main beam and side lobes of an elevation-scanned, high-gain, fan beam receiving antenna. The satellite is assumed to be flying a course directly over the antenna, along the direction of antenna scan.

Analysis

The radiation pattern of a line source of length 2a along the x-direction is proportional to

$$g(\phi) = \int_{-1}^{1} E(x/a) \exp -j\left[\phi x/a - \gamma(x/a)\right] d(x/a) \qquad (1)$$

where

- $E(x/a)$ = magnitude of the aperture distribution
- $\gamma(x/a)$ = phase of the aperture distribution
- ϕ = $2\pi a/\lambda \sin\theta$
- θ = angle between the direction considered and the normal to the surface.

For the case of a sinusoidal phase error we let $\gamma(x/a)$ be of the form $\beta\cos(n\pi x/a + \eta)$ where η is a constant controlling the position of maximum error, and $E(x/a)$ be $E_0(x/a)$, so that Eq. (1) becomes

$$g(\phi) = \sum_{m=0}^{\infty} \frac{j^m \beta^m}{m!} \int_{-1}^{1} E_0(x/a) \cos^m(n\pi x/a + \eta) \exp(-j\phi x/a)\, d(x/a). \qquad (2)$$

From the theory of Fourier Integrals the transform of a product of two functions is the convolution of their transforms.[5] Eq. (2) then can be written as

$$g(\phi) = \sum_{m=0}^{\infty} \frac{j^m \beta^m}{m!} \left\{ g_0(\phi) * \frac{1}{2}\left[\delta(\phi+n\pi) + \delta(\phi-n\pi)\right] \right\} \qquad (3)$$

where

$$g_0(\phi) = \int_{-1}^{1} E_0(x/a) \exp(-j\phi x/a)\, d(x/a)$$

- = radiation pattern with no phase error
- β = peak phase error across the aperture
- δ = impulse function
- n = number of cycles of error along the line source
- η = set equal to zero

$$g_0(\phi) * h_0(\phi) = \left[g_0(\phi) * h_0(\phi)\right] * h_0(\phi)$$

and the asterisk represents convolution.

For a sinusoidal amplitude error, $E(x/a)$ is of the form $E_0(x/a)[1 + b\cos(n\pi x/a + \epsilon)]$ where ϵ is a constant controlling the position of maximum error. Eq. (1) then becomes

$$g(\phi) = \int_{-1}^{1} E_0(x/a)\left[1 + b\cos(n\pi x/a + \epsilon)\right] \exp(-j\phi x/a)\, d(x/a). \qquad (4)$$

From the convolution property of Fourier transforms this becomes

$$g(\phi) = g_0(\phi) - b\left\{ g_0(\phi) * \frac{1}{2}\left[\delta(\phi+n\pi) - \delta(\phi-n\pi)\right] \right\} \qquad (5)$$

where

- b = peak amplitude error across the aperture
- ϵ = set equal to $\pi/2$.

Finally, for a quadratic phase error we let $\gamma(x/a)$ be of the form $-\beta(x/a)^2$, and $E(x/a)$ be $E_0(x/a)$. Eq. (1) then becomes

$$g(\phi) = \sum_{m=0}^{\infty} \frac{(-j)^m \beta^m}{m!} \int_{-1}^{1} (x/a)^{2m} E_0(x/a) \exp(-j\phi x/a)\, d(x/a) \qquad (6)$$

* The work reported in this paper was supported by the Ballistic Research Laboratories under Contract DA-04-200 ORD-874.

From the theory of Fourier Integrals the transform of a product of "x/a" and a function f(x/a) is "j" times the derivative of the transform of f(x/a).[3] Thus Eq. (6) can be written as

$$g(\phi) = \sum_{n=0}^{\infty} \frac{j^n \beta^n}{n!} \frac{d^{2n}}{d\phi^{2n}} [g_0(\phi)] \qquad (7)$$

where β = error at the edge of the aperture.

If we separate $g(\phi)$ into its real and imaginary parts

$$g(\phi) = R + jI . \qquad (8)$$

Then for Eq. (3) the real and imaginary parts for a cosinusoidal phase error are

$$R = g_0(\phi)[1-\beta^2/4] - \beta^2/8[g_0(\phi+2n\pi) + g_0(\phi-2n\pi)] + \ldots$$

$$I = \beta/2[1-\beta^2/8][g_0(\phi+n\pi) + g_0(\phi-n\pi)]$$

$$-\beta^3/48[g_0(\phi+3n\pi) + g_0(\phi-3n\pi)] + \ldots$$

where $g_0(\phi+n\pi)$ and $g_0(\phi-n\pi)$ are reproductions of the radiation pattern with no phase error, and located symmetrically on either side of the main lobe at an angle of $n\pi$ radians from it.

For a sinusoidal amplitude error the real and imaginary parts are

$$R = g_0(\phi)$$

$$I = b/2[g_0(\phi-n\pi) - g_0(\phi+n\pi)]$$

while for a quadratic phase error they become

$$R = g_0(\phi) - \frac{\beta^2}{2} \frac{d^4}{d\phi^4} [g_0(\phi)] + \ldots$$

$$I = \beta \frac{d^2}{d\phi^2} [g_0(\phi)] - \frac{\beta^3}{6} \frac{d^6}{d\phi^6} [g_0(\phi)] + \ldots$$

For any of these types of error the phase angle at constant radius from the center of the aperture is then

$$\psi(\phi) = \tan^{-1} I/R \quad \text{degrees} \qquad (9)$$

and the magnitude of the radiation pattern is

$$|g(\phi)| = \sqrt{R^2 + I^2} . \qquad (10)$$

The instantaneous frequency associated with the phase angle is

$$f = \frac{1}{2\pi} \frac{d\psi(\phi)}{dt} \quad \text{cycles/second} \qquad (11)$$

or

$$f = \frac{1}{\lambda} \left(\frac{d\theta}{dt}\right) \frac{d\psi(\phi)}{d\theta} \quad \text{cycles/second} \qquad (12)$$

where

$$\frac{d\theta}{dt} = v/R \pm \frac{dx}{dt}$$

and $\sin\theta \approx \theta$.

The plus sign is chosen if the satellite velocity is in a direction opposite to the beam scanning direction.

v = satellite velocity
R = range from satellite to receiver
dx/dt = angular scanning rate of the beam.

In a similar manner, the instantaneous acceleration associated with the phase angle is

$$A = df/dt \quad \text{cycles/second/second} \qquad (13)$$

or

$$A = 2\pi \left(\frac{\theta}{\lambda} \frac{d\theta}{dt}\right)^2 \frac{d^2\psi(\phi)}{d\phi^2} \quad \text{cycles/second/second.} \qquad (14)$$

Discussion

Figures 1 through 8 illustrate the effects, for a uniformly illuminated aperture (E_0 = constant), of cosinusoidal variations in phase, sinusoidal variations in amplitude, and quadratic variations in phase across the aperture, on the phase angle, radiation pattern, instantaneous frequency, and acceleration, for values of β = 0, $\pi/8$, $\pi/2$, b = 0, $\pi/8$ and π = 1. To keep the figures in the most general form, curves of $d\psi/d\theta$ and $2\pi d^2\psi/d\theta^2$ instead of frequency and acceleration have been plotted respectively in Figs. 5 and 6 and in Figs. 7 and 8. In addition, all of the figures have been plotted against ϕ. However, for large (in wavelengths) apertures this angle is directly proportional to the actual angle θ in the vicinity of the main beam. Choosing the period of the error equal to the aperture width locates the largest error pattern $[g_0(\phi \pm n\pi)]$ at the first nulls of the original pattern $[g_0(\phi)]$. The region of the pattern over which computations were made lies between the peak of the pattern (ϕ = 0) and the part halfway between the first and second null ($\phi \approx \pm 270$). The computed values are based on calculations using the first four terms of the series

in Eqs. (3) and (4) for the phase error, while the exact expression is used for the amplitude error.

Figures 1 and 2 show that for a uniform amplitude distribution without any phase or amplitude error ($\beta = 0$, $b = 0$) the phase angle has a step discontinuity of 180 degrees as θ passes from the main beam to the side lobe. As phase error is introduced, the variation in phase angle along the main beam and the first side lobe becomes more gradual and the total phase variation over the region considered becomes smaller. As sinusoidal amplitude error is introduced, the variation in phase angle is similar to the above, with the exception that the resultant curve is an odd function about the direction of the peak, whereas the curves for the phase error are even functions. As a result, one would expect the instantaneous frequency and acceleration at the peak of the pattern to be zero and finite, respectively, for the phase error, and finite and zero, respectively, for the amplitude error.

As shown in Figs. 3 and 4, the general effect of the errors on the radiation pattern is to raise both the side-lobe level and the level of the minima with a corresponding loss in maximum gain.

In Fig. 3 it is seen that with $\beta = \pi/2$ for the cosinusoidal phase error the main lobe has become bifurcated with maxima appearing on either side of the $\theta = 0$ axis; the side-lobe level has been raised 8 db above the level calculated with no phase error; the minimum has been raised from a complete null to only 9 db below the peak; and the gain in the $\theta = 0$ direction is down 8 db from that calculated with $\beta = 0$. Correspondingly, with $\beta = \pi/2$ for the quadratic phase error the side-lobe level has been raised 3 db above the level calculated with no phase error, the minimum has been raised from a complete null to only 9.8 db below the peak, and the maximum gain is down 1 db from that calculated with $\beta = 0$.

Since the frequency error is proportional to the slope of the phase-angle variation, Figs. 5 and 6 are logical extensions of Figs. 1 and 2. The step discontinuity in the phase angle found with no error ($\beta = 0$, $b = 0$) produces a delta-function frequency error at the times corresponding to the satellite passage at the first null of the radiation pattern. However, when either a phase or an amplitude error is introduced, the resultant smoothing of the phase-angle variation results in finite frequency errors. The peak of the frequency variation is still in the vicinity of the null, but shifts toward the peak of the pattern as the phase error is increased; the maximum frequency error becomes smaller, and the variation occurs over a larger time interval. Therefore, the frequency error at the half-power points of the pattern increases as the phase error is increased. As mentioned before, a non-zero frequency error exists on the $\theta = 0$ axis for the amplitude-error conditions and the resultant curve is now an even function about this axis while the curves for phase errors are odd functions.

Since the acceleration error is proportional to the slope of the frequency-error variation, the discussions used for explaining the frequency-error variation will also be applicable for explaining the acceleration-error variation shown in Figs. 7 and 8. Since the slope of the frequency-variation changes sign, the acceleration error also changes sign.

Effect on Satellite Tracking

To obtain the amount of Doppler frequency shift resulting from small phase errors along the aperture ($\beta \leq \pi/6$) it is necessary to multiply the value of $d\phi/d\theta$ obtained from Figs. 5 and 6 by $\frac{a}{\lambda}\frac{d\theta}{dt}$. This operation combines the rate of change of phase angle with respect to the angular direction and the rate of change of angular direction with respect to time producing the rate of change of phase angle with time (frequency). To obtain the acceleration the value of $2\pi\frac{d^2\phi}{d\theta^2}$ obtained from Figs. 7 and 8 are multiplied by $\left(\frac{a}{\lambda}\frac{d\theta}{dt}\right)^2$. Typical antenna parameters of the BRL Doploc system* are a one degree beamwidth fan beam with a scanning rate $\left(\frac{d\alpha}{dt}\right)$ of 18 degrees per second. With these values a/λ is approximately 26 and $d\theta/dt$ becomes .28 radians per second for a 300-mile-altitude satellite travelling with a 8-mile-per-second velocity. For these values a typical frequency shift within the half-power beamwidth due to a cosinusoidal phase error which varies from $\pi/6$ radians at the center of the aperture to $-\pi/6$ radians at the edges of the aperture is about 2 cycles per second while that due to a quadratic phase error which varies from zero radians at the center to $-\pi/6$ radians at the edges of the aperture is about one cycle per second. The respective accelerations introduced are about 150 cycles per second per second and 60 cycles per second per second; a peak amplitude error of approximately one-half the original aperture amplitude would be required to give frequency and acceleration errors comparable to those due to small phase errors. Also from the calculated data it can be seen that comparable errors are produced for a peak cosinusoidal error which is 1/2.5 as large as the quadratic phase error at the edge of the aperture. This relation seems reasonable since the first Fourier coefficient in the expansion of a quadratic function is $4/\pi^2$.

---

* The Doploc system is a bistatic, CW, reflection Doppler-detection system concerned with the detection of non-radiating satellites and the determination of their orbits. The orbit of the satellite can be determined from the Doppler shifted received signals reflected from the satellite.

Since the information bandwidths required to give adequate signal-to-noise ratios to maintain lock are proportional to the square root of the rate of change of the Doppler frequency, the effect of an additional acceleration will be to increase the required bandwidths. For satellites in orbit at an altitude of 100 miles, the required bandwidth increases from 10 cps to 16 cps, while for satellites in orbit at a 2,000-mile altitude, the required bandwidth increases from 3 to 10 cps. Therefore, a heavy penalty is paid in terms of total transmitter power required by the increase in rate of change of Doppler shift.

By examining Eq. (9) one may see the effect of cosinusoidal phase errors with several cycles of error along the line source on the phase angle and hence on the frequency and acceleration errors. For $(\beta^2/4) << 1$ the phase-angle error is approximately

$$\psi(\phi) \cong \tan^{-1} \frac{\beta}{2} \frac{[g_0(\phi+n\pi) + g_0(\phi-n\pi)]}{g_0(\phi)} \text{ degrees.} \tag{15}$$

For a uniformly illuminated aperture

$$g_0(\phi') = \frac{\sin \phi'}{\phi'} \tag{16}$$

Substituting Eq. (16) into Eq. (15),

$$\psi(\phi) \cong \tan^{-1} (-1)^{n-1} \frac{\beta}{\left(\frac{n\pi}{\phi}\right)^2 - 1} \text{ degrees} \tag{17}$$

for $n > 1$.

Between the half-power points

$$\psi(\phi) \cong \frac{(-1)^{n-1}}{n^2} \left(\frac{\beta}{2}\right) \phi^2 \text{ degrees.} \tag{18}$$

From Eqs. (12) and (14) we see that

$$f \cong \frac{(-1)^{n+1} 2}{n^2} \frac{a}{\lambda} \left(\frac{d\theta}{dt}\right)\left(\frac{\beta}{2}\right) \phi \left[1 + 2\left(\frac{\phi}{n\pi}\right)^2\right] \tag{19}$$

and

$$A \cong (-1)^{n+1} \frac{4}{\pi} \left(\frac{a}{\lambda} \frac{d\theta}{dt}\right)^2 \left(\frac{\beta}{2}\right) [1 + 8(\phi/n\pi)^2] . \tag{20}$$

From these equations it is seen that as the frequency of the phase-error distribution increases, the peak phase error decreases by a factor of $1/n^2$ for angular directions between the half-power points. Consequently, only the lower-harmonic sinusoidal phase errors contribute any appreciable frequency or acceleration error.

Multiple Aperture Error Variations

If we now consider any small phase-error distribution across the aperture as consisting of a sum of sinusoidal distributions for various numbers of cycles of error then

$$\gamma(x/a) = \sum_{n=1}^{N} \beta_n \cos(n\pi x/a). \tag{21}$$

Substituting into Eqs. (1) and (2) the resultant radiation pattern is proportional to

$$g(\phi) = \sum_{m=0}^{\infty} \frac{j^m}{m!} \left\{ g_0(\phi)^m \frac{1}{2} \sum_{n=1}^{N} \beta_n^m [\delta(\phi+n\pi) + \delta(\phi-n\pi)] \right\}. \tag{22}$$

For $\beta_n << 1$ Eq. (22) simplifies to

$$g(\phi) = g_0(\phi) + \frac{j}{2} g_0(\phi) * \sum_{n=1}^{N} \beta_n [\delta(\phi+n\pi) + \delta(\phi-n\pi)]. \tag{23}$$

If we separate $g(\phi)$ into its real and imaginary parts

$$R = g_0(\phi)$$

$$I = \frac{1}{2} \sum_{n=1}^{N} \beta_n [g_0(\phi+n\pi) + g_0(\phi-n\pi)] .$$

The phase angle for $\psi \simeq \tan \psi$ at constant radius from the center of the aperture is then approximately

$$\psi(\phi) \cong \frac{1}{2R} \sum_{n=1}^{N} \beta_n [g_0(\phi+n\pi) + g_0(\phi-n\pi)]. \tag{24}$$

Since the equation and the subsequent equations for frequency and acceleration are linear, the superposition theorem can be used to determine the effects of a sum of N cosinusoidal phase error components. It can also be demonstrated that a similar relation also holds for a sum of sinusoidal amplitude errors. Since the equation for the phase angle is not linear for values of ϕ such that $\phi \not\approx \tan \psi$, the frequency and acceleration errors can only be obtained by considering the complete expression for I instead of by the superposition of the individual effects of its component parts.

Effects due to Phase Center

In the preceding discussion we have considered that the phase center of the antenna and the point of rotation are coincident. By examining Eqs. (13) and (14) one may determine the effect of a displacement between the phase center and the point of rotation on the instantaneous frequency and acceleration. If the phase center lies behind the rotation point a distance "d" then the corrected phase angle, ϕ_c, is

$$\phi_c(\theta) = \phi(\theta) + kd\cos\theta \quad \text{radians} \qquad (24)$$

where

$\phi(\theta) = \tan^{-1} I/R$ = phase angle for d zero
I = imaginary part of radiation pattern
R = real part of radiation pattern
$k = 2\pi/\lambda$
θ = angle between the direction considered and the normal to the surface.

From Eqs. (12) and (25) we see that the corrected instantaneous frequency is

$$f_c(\theta) = f(\theta) - \frac{d}{\lambda}\frac{d\phi}{dt}\sin\frac{\lambda\theta}{2\pi a} \quad \text{cps} \qquad (26)$$

and the corrected instantaneous acceleration is

$$A_c(\theta) = A(\theta) - \frac{d}{\lambda}\left(\frac{d\phi}{dt}\right)^2\cos\frac{\lambda\theta}{2\pi a} \quad \text{cps/sec.} \qquad (27)$$

where

$\theta = 2\pi a/\lambda \sin\theta$
$\sin\theta \approx \theta$
$f(\theta)$ = instantaneous frequency with d zero
$A(\theta)$ = instantaneous acceleration with d zero

In the vicinity of the main beam these equations reduce to

$$f_c(\theta) = f(\theta) - \frac{d}{\lambda}\theta\left(\frac{d\phi}{dt}\right) \quad \text{cps} \qquad (28)$$

$$A_c(\theta) = A(\theta) - \frac{d}{\lambda}\left(\frac{d\phi}{dt}\right)^2 \quad \text{cps/sec.} \qquad (29)$$

Therefore, the effect of a non-coincident phase center and point of rotation is to introduce an additional component frequency and acceleration error into the received signal. This may be used to reduce the on-axis acceleration error.

Conclusions

An analysis has been carried out of the effect of the antenna phase pattern on the instantaneous frequency shift and the rate of change of frequency shift of a received Doppler signal as a satellite passes through the side lobes and main beam of a scanning narrow fan beam antenna. Approximate formulas and graphs are presented which illustrate these effects for a uniformly illuminated aperture with either small cosinusoidal variations in phase, or sinusoidal variations in amplitude, or quadratic variations in phase across the aperture. Although the analysis only considers the effects of cosinusoidal and quadratic phase errors and sinusoidal amplitude errors for practicality, it could be extended to consider the general case of even functions of phase error and odd functions of amplitude error. Since odd functions of phase error or even functions of amplitude error do not introduce a quadratic component into the radiation pattern, these errors by themselves will not affect the phase pattern. If one were to consider a phase and amplitude error existing simultaneously across the aperture, both even and odd functions would have to be considered.

An example of a one-degree-beamwidth fan beam antenna scanned at a rate of 15 degrees per second shows that between the half-power points the maximum Doppler frequency shift is about one cycle per second while the maximum rate of change of frequency is about 100 cycles per second per second. If the Doppler frequency introduced by the aperture errors is an odd function about the peak of the pattern ($\theta = 0$), the frequency shift may be removed from the received signal if the detector is capable of averaging the frequency variations over a time interval. Since the information bandwidth required in a phase-lock receiver is proportional to the square root of the rate of change of the Doppler frequency, this effect may require an appreciable increase in the minimum allowable bandwidth with a subsequent reduction in the signal-to-noise ratio.

References

1. L. G. de Bey, V. W. Richard, R. B. Patton, "Second Semi-Annual Tech. Summary Report Period January - 30 June 1959," NRL Memorandum Report No. 1220 (July 1959) CONFIDENTIAL.
2. H. L. Richter, W. F. Sampson, R. Stevens, "Microlock: A Minimum Weight Radio Instrumentation System for a Satellite," Jet Propulsion 28, pp. 532-540 (August 1958).
3. C. L. Nielsen, "Principles and Applications of Phase-Lock Detection in Phase-Coherent Systems," Jet Propulsion 28, pp. 841-847 (August 1958).
4. V. W. Richard, "Duploc Tracking Filter," NRL Memorandum Report No. 1173 (October 1958).
5. I. N. Sneddon, Fourier Transforms (McGraw-Hill Book Company, Inc., New York, N.Y., 1951).

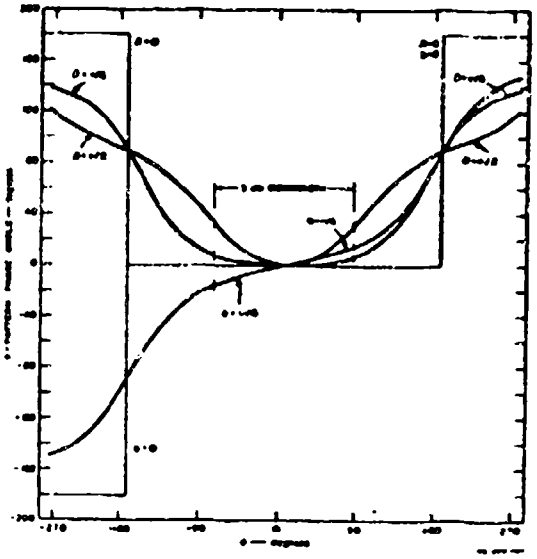


Fig. 1

The effect of Introducing a Cosinusoidal Phase or Sinusoidal Amplitude Error on the Radiation Pattern Phase Angle

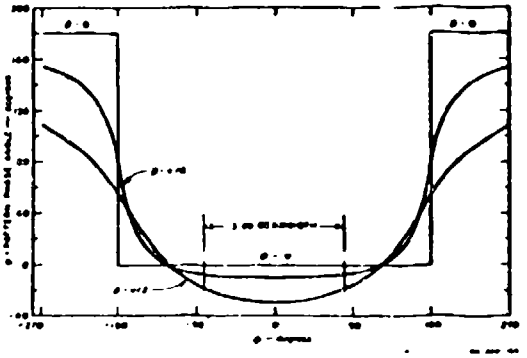


Fig. 2

The Effect of Introducing a Quadratic Phase Error on the Radiation Pattern Phase Angle

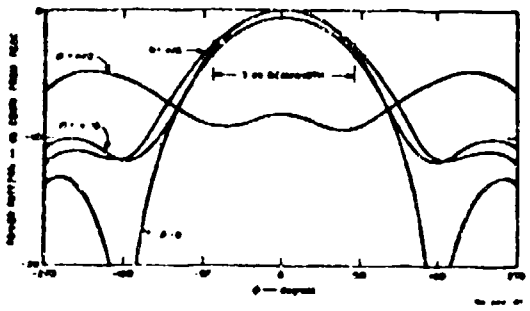


Fig. 3

The Effect of Introducing a Cosinusoidal Phase or Sinusoidal Amplitude Error on the Radiation Pattern Magnitude

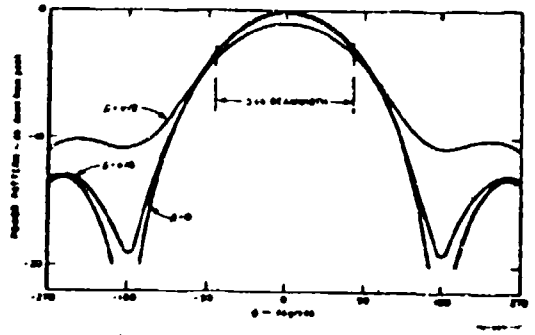


Fig. 4

The Effect of Introducing a Quadratic Phase Error on the Radiation Pattern Magnitude

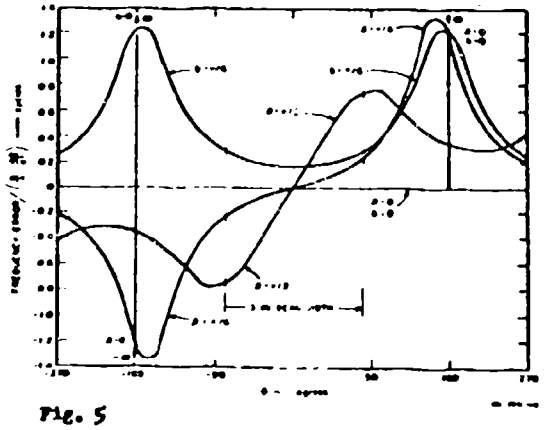


Fig. 5

Frequency Error Introduced by a Cosinusoidal Phase or Sinusoidal Amplitude Error

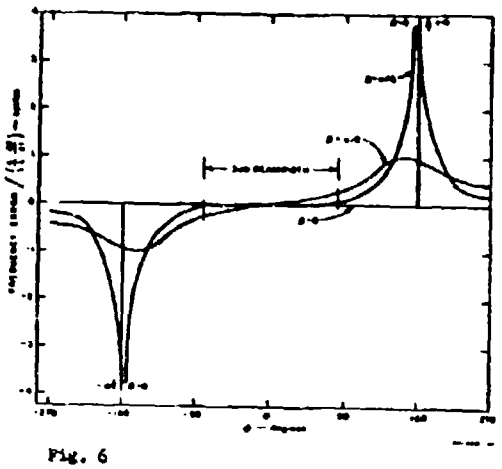


Fig. 6

Frequency Error Introduced by a Quadratic Phase Error

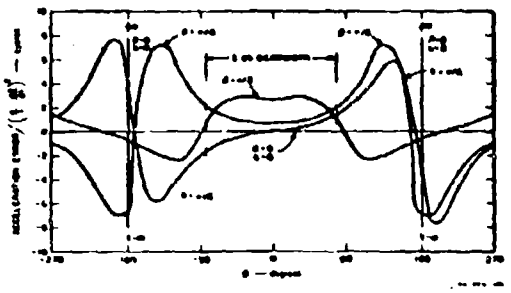


Fig. 7

Acceleration Error Introduced by a Cosinusoidal Phase or Sinusoidal Amplitude Error

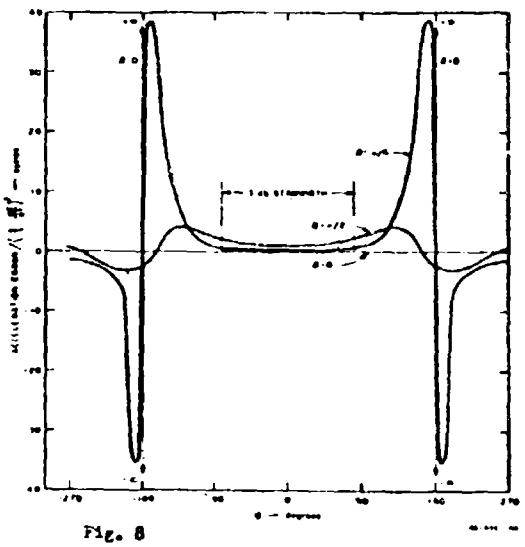


Fig. 8

Acceleration Error Introduced by a Quadratic Phase Error

THE SHROUD ANTENNA

By: Mr. Victor W. Richard, Ballistic Measurements Laboratory
Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

ABSTRACT

The Shroud antenna was designed for VHF telemetering and Doppler tracking instrumentation use on high velocity rockets which require a small, lightweight antenna with an omnidirectional radiation amplitude, constant phase, pattern. A unique method of installation of the Shroud antenna system is used which does not require rocket body modification, mounting holes or bolts. The antenna elements are held fast to the rocket body with metal straps, facilitating convenient, secure mounting. The Shroud antenna has withstood successfully the high acceleration and temperature environment of the Nike-Cajun rocket which has a peak acceleration of 80 g's and attains a speed of mach 6 as low as 50,000 feet. The flat plate area of the antenna is small, presenting low drag. High temperature resistant materials are used to withstand the extreme heat generated by air friction.

Properties of the Shroud antenna discussed include: the basic theory of operation, radiation amplitude and phase patterns, gain, efficiency, voltage breakdown, radiation efficiency, impedance matching, construction installation, adjustment and flight performance.

INTRODUCTION

The Shroud antenna was designed to fill the need for a special rocket antenna system in connection with the Upper Atmosphere Research Program of the Ballistic Research Laboratories and the ballistic instrumentation of rockets fired at the Fort Churchill ECY rocket range during the recent International Geophysical Year program. The DOVAP (DOppler Velocity And Position) tracking and telemetering system used required a rocket antenna system which had known phase properties and provided a continuous amplitude of the signal transmitted from the ground and for retransmission back to the ground. Reliable operation was required under the extreme environment of the small, high velocity rockets with high take-off acceleration and high velocity at low altitudes. The method of installation was required to be simple and convenient with little or no modification of the rocket body permitted since space available for the antennas was on the outer shell of the rocket which contained the fuel.

With few exception, the Shroud antenna system met these conditions and was used on the 60 sounding rockets fired for the IGY program[1,2] that carried DOVAP instrumentation.

SHROUD ANTENNA DESCRIPTION

General Description

The Shroud antenna is basically a loop antenna. The loop is rectangular in form, electrically balanced, resonant at the operating frequency, and matched to 50 ohms. The name "Shroud" was derived from its similarity in cross section and appearance to the cable cover shrouds used on rockets. Figure 1 shows the shroud shaped antenna configuration as installed on the Cajun rocket. These antennas were used with the DOVAP tracking and telemetering system for reception at 58 mc and transmission at 76 mc. Four Shroud elements were employed for this use; two for reception and two for transmission.

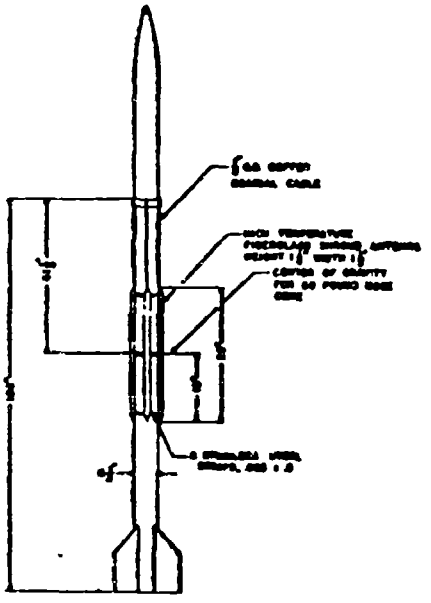


Figure 1. Shroud Antennas On Cajun Rocket

Shroud-Loop Development

The continuous information sought with telemetering and ballistic tracking throughout the entire rocket flight, from launching to impact, imposed a requirement for the antenna system to be as close to isotropic as possible. The nearly vertical firings above the ground tracking stations dictated good radiation coverage downward from the tail region of the rocket. Thus, nose stub or body fed antennas were not usable since they would have a deep null off the rocket tail. The basic requirement for radiation in the direction of the longitudinal axis of the rocket is to set up an electric field perpendicular to this axis. A loop antenna whose plane contains the longitudinal axis of the rocket fills this requirement. Studies conducted at the Ballistic Research Laboratories over the past ten years on VHF (30 to 300 mc) antennas for small rockets

have shown that the loop has many desirable properties and advantages over dipole and slot configurations. Slots, which must be transverse for radiation along the rocket axis, are much too short for efficient radiation when the rocket diameter is only of the order of 6 inches, (1/50 λ at 38 mc). Dipoles and stubs must protrude from each side of the rocket body by at least λ/10 (2.5 feet at 38 mc) for acceptable radiation efficiency. This is usually an excessive protrusion for reasons of drag, stability and mechanical design considerations.

A free space circular loop, whose circumference is less than one-half wavelength, has an omnidirectional pattern in the plane of the loop. Experimental studies have disclosed that the shape of the loop has no appreciable effect on the shape of the radiation pattern as long as the circumference is less than one-half wavelength, i.e., there is a unidirectional current flow around the loop with no phase reversal. The upper part of Figure 2 shows two rectangular loop shapes of proportions which still yield an omnidirectional radiation pattern. An early application of the balanced loop antenna for DOVAP tracking and telemetering on the two stage WAC Corporal-V2 rocket is shown in Figure 2. The measured radiation pattern is seen to be only slightly perturbed by the long, large rocket body. The loop was mounted in the nose section and a fiberglas nose cone cover was used. As space became unavailable for antennas in subsequent rocket experiments, it became necessary to move the antennas to the side of the body. Figure 2 shows two rectangular loops mounted on an Aerobee rocket. Two loops are used to form a balanced system to eliminate the effect of the rocket body on the radiation pattern. The two loops, which are fed 180° out of phase, excite equal amplitude, out of phase currents along the rocket body. Thus the longitudinal body current, which would radiate in the dipole mode and cause a deeply lobed radiation pattern, are cancelled. The loop current direction of flow depicted in Figure 2 on the Aerobee rocket shows the opposite directions of current flow along the rocket body. By use of this technique the perturbing effect of the rocket body on the radiation pattern is effectively made negligible as illustrated by the measured radiation pattern in Figure 2. The two loop array offers the very desirable advantage of permitting the antenna installation anywhere along the rocket body as required by aerodynamic or other considerations. Complete measured radiation amplitude patterns of a two loop array on an Aerobee rocket are shown in Figure 3, demonstrating the close agreement obtained in practice with a free space loop. Measured radiation phase patterns are given in Figures 4 and 5. The radiation phase around the longitudinal axis of the rocket is seen to be constant within 10°. In the transverse plane the expected 180° abrupt phase shift is obtained. The phase modulation caused by rocket spin that an antenna with these phase properties will impose on the carrier has been studied by the Stanford Research Institute under Contract DA-04-200-ORD-273 with the Ballistic Research Laboratories, and is reported in References 3, 4, 5 and 6.

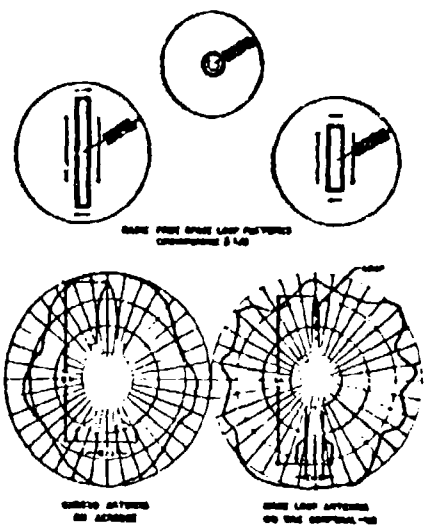


Figure 2. Loop Antenna Radiation Patterns of Free Space and Rocket Mounted Loops

Shroud Construction

The rectangular loop configuration of the shroud antenna is shown in Figure 6. The conductor which passes along the length and ends of the shroud forms three sides of the loop. The surface of the rocket body serves as the fourth side, completing the electrical path for the loop circulating current. The metal straps at the ends press the end feet tightly against the rocket body for electrical contact and physical support. If a metallic outer surface is not available, a narrow strip of metal, foil, or conducting paint may be run along the body.

The loop is electrically balanced with respect to the rocket body with a gap at its mid-length. The impedance at the gap consists of a resistive component, of the order of 50 ohms, in series with a large inductive reactance. The reactance is inductive by design since the circumference of the loop is made to be less than one-half wavelength. This allows resonating the antenna with a small variable capacitor which has low loss and is mechanically convenient. The matching circuit employs inductive coupling between the feed cable and an inductor in series with the resonating capacitor at the gap. An adjustment of the input resistive component is provided by the slug in the coupling coil. The variable capacitor is used to adjust for zero input reactance and for a further fine phase adjustment. The variable capacitor is shunted with a fixed capacitor to provide vernier tuning control and temperature compensation. The fixed capacitor is chosen to have a change of capacitance with temperature equal and opposite to that of the variable capacitor. By reversing the leads

to the coupling transformer from the coaxial cable (A and B, Figure 6) the radiation phase can be changed 180°. This is done when using a two loop array to provide the required 180° phase difference between them. The coaxial cable commonly used has Teflon insulation and a solid copper outer sheath to permit operation at high temperatures and for physical rigidity. For extremely high temperature operation, a commercially available coaxial cable with magnesium oxide dielectric insulation can be used.

An assembly drawing of the molded Fiberglas shroud antenna is shown in Figure 7. A method of molded construction is used to facilitate quantity production. The outer shell is formed of several layers of No. 112 weave Fiberglas cloth impregnated with 91 LD phenolic resin. (91 LD resin is a product of the Cincinnati Testing Laboratories, 316 W. Fourth Street, Cincinnati, Ohio). To form the shroud, the first layer of Fiberglas cloth is placed directly on the bottom half of the mold. A copper foil conductor is applied next and then additional layers of Fiberglas cloth are placed over the foil conductor until the required thickness is obtained. The top half of the mold is placed in position and the entire assembly is subjected to high pressure and temperature until the resin is cured. The 91 LD resin is a phenolic resin that has very high short time temperature resistant properties. The ends of the shroud antenna are made thicker than the sides for increased temperature resistance. They will withstand 1500°F for about 15 seconds. Shrouds recovered from Cajun flights showed only surface discoloration with no surface damage from the heat. Measurements made in flight show temperatures up to 1800°F on the surface of the leading edge of the shroud and 400°F in the gap matching section. The shroud Fiberglas wall is also

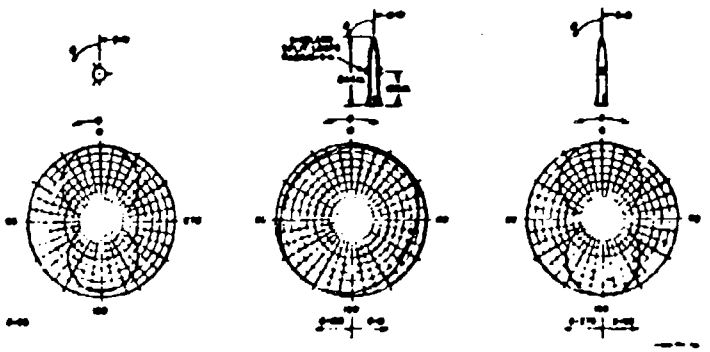


Figure 3. Amplitude Patterns of Loop Antenna on Aerobee Missile---37 Mc.

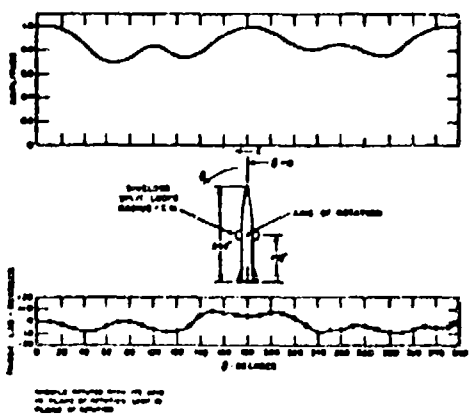


Figure 4. Phase and Amplitude Patterns of Loop Antenna on Aerobee Missile--74 Mc.

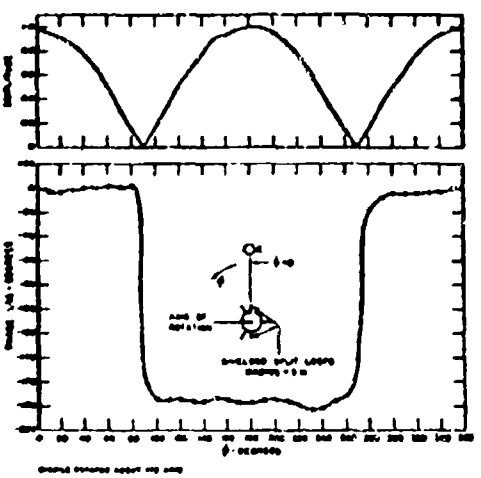


Figure 5. Phase and Amplitude Patterns of Loop Antenna on Aerobee Missile--74 Mc.

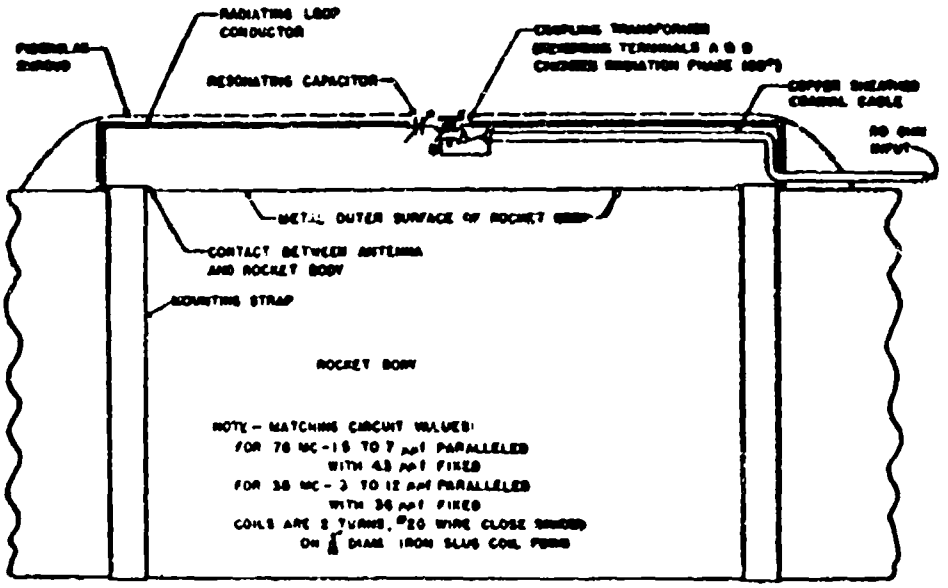


Figure 6. Circuit Diagram of Shroud Antenna

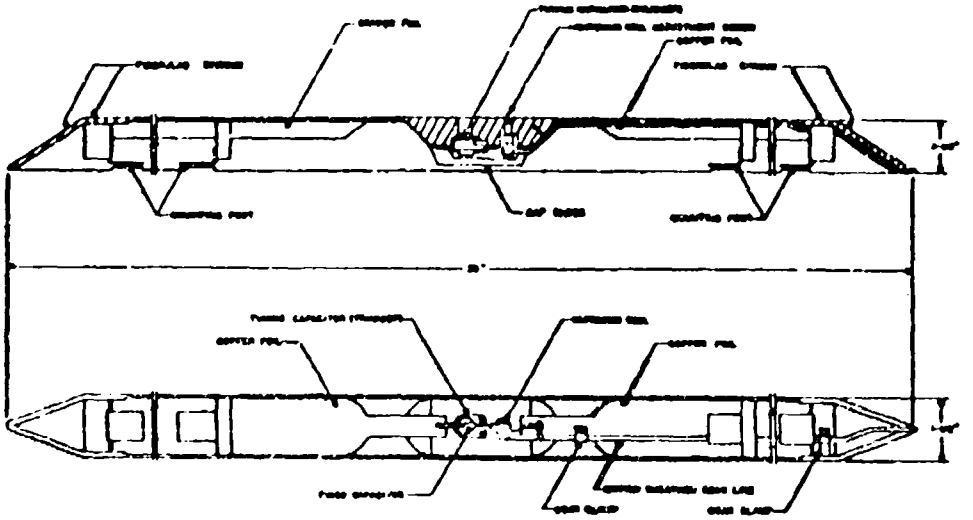


Figure 7. Assembly Diagram of Fiberglass Shroud Antenna

made thicker at the antenna gap to provide thermal insulation for the tuning capacitor. The matching components at the gap are completely enclosed by a Fiberglas block and a silicone compound, DC-4, is forced in to fill the voids for voltage breakdown protection.

SHROUD ANTENNA INSTALLATION AND ADJUSTMENT

A complete Shroud antenna installation on a Cajun rocket is shown in Figure 8. The additional straps between the end straps pull the antenna tightly to the rocket body to prevent vibration during flight. The tapered blocks in front of the antennas serve as convenient disconnect junctions between the coaxial cable from the antenna and the instrument compartment. They also provide some protection for the leading edge of the antenna by creating a shock wave pattern which diverts some of the high velocity air away from the antenna.

equal in magnitude and opposite in phase. This near field measurement has proven to be a reliable method of insuring the proper far field radiation pattern. It has the very practical advantage of permitting antenna adjustments within the close proximity of the ground and large metallic objects. Special matching and phasing test equipment have been developed for Shroud antenna tuning. Figure 9 shows the test setup and antenna test equipment.

All antennas are adjusted as transmitting antennas using a directional coupler for VSWR indication and a low power test signal source of about 1 watt power level. Reference 7 gives additional details on the test equipment and the tuning procedures.

The installation of Shroud antennas on a Spaerobee rocket is shown in Figure 10. The antennas in this case were mounted quite close

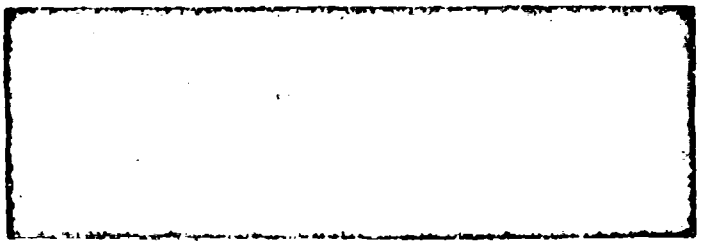


Figure 8. Shroud Antenna Installation on Cajun Rocket

The Shrouds are connected in pairs, diametrically opposed, and fed with equal length coaxial lines which are paralleled in the instrument compartment. The resultant parallel impedance of the two 50 ohm antennas (25 ohms) is transformed up to 50 ohms with an LC matching network. In practice the VSWR can be adjusted to better than 1.2:1.

The procedure for matching and phasing is to adjust each antenna individually for the lowest VSWR (when fed separately they are not designed to be exactly 50 ohms because of the allowance made for mutual coupling). The antennas are then fed as a pair and readjusted to compensate for the effects of mutual coupling. The final tuning procedure consists of making adjustments of the variable capacitors on each antenna until a low VSWR is obtained coincident with a 180° phase difference between them. A coarse adjustment of the tuning capacitor is sufficient to bring each loop into resonance and give a low VSWR, then further fine adjustment is possible to get the proper phase without increasing the VSWR appreciably. The phase difference between the antennas is measured by placing a field strength indicator at an equal distance from each antenna and about 8 inches from the surface of the rocket body. The field strength at this point will be zero when the radiation from each antenna is

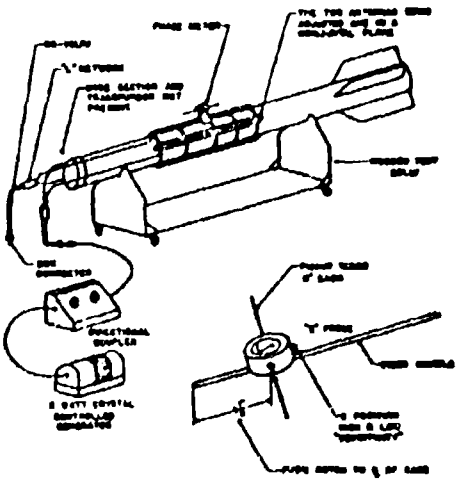


Figure 9. Matching and Phasing Set-up

to the tapered section of the rocket nozzle and also had to be unevenly spaced around the rocket body to fit along with the cable shrouds. The radiation pattern and impedance of the antenna system was not appreciably affected by this arrangement.

The installation of Shroud antennas on an Aerobee-Hi rocket is shown in Figure 11. This figure illustrates the very small size of the Shroud antenna relative to the Aerobee rocket dimensions and cable shrouds. The shrouds shown in Figure 11 are an early version with a triangular cross section and a stainless steel outer shell.

The weight of a complete Shroud antenna installation, including straps and coaxial cable is about 5 pounds. Its maximum protrusion from the rocket body is 1 1/2 inches.

FLIGHT PERFORMANCE

The Shroud rocket antenna system has proven capable of providing excellent continuous signals of fine quality at the ground stations. By using circularly polarized receiving antennas, completely continuous, high signal-to-noise, telemetering and DOVAP records were consistently obtained from a transponder with 1 1/4 watts average power output at altitudes up to 150 miles.

The VSWR of the Shroud transmitting antenna system has been measured during flight by telemetering the output of a directional coupler installed in the transmitting antenna feed cable. The telemetered data has shown an increase in VSWR during some of the flights in the final phase of maximum motor thrust, probably due to thermal effects from the aerodynamic heating of the antenna and the antenna tuning elements. A second region of VSWR increase is noted between 25 and 75 miles altitude. This is caused by the ionization of the air in the vicinity of the transmitting antenna gap. The received ground signal diminishes during voltage breakdown of the gap by as much as 12 db but a minimum of 5 microvolts has always been received by the ground receiver which is sufficient for proper ground station operation. One watt input power to the transmitting antenna results in a gap voltage of approximately 90 volts. This large gap voltage is a natural consequence of the small size and high Q of the antenna. When a small antenna is designed for efficient radiation, the gap voltage will be high. The field gradient in the vicinity of the gap is large enough to cause ionization of the rarified air encountered at altitudes between 25 and 75 miles. Complete prevention of ionization with 1 watt input to the antenna system would require about 6 inches of insulation around the gap which would make the flat plate area and drag excessively large for most high velocity rockets.

The pattern shape of the two Shroud antenna system is that of a doublet antenna, therefore, its pattern gain over isotropic is 1.76 db. The

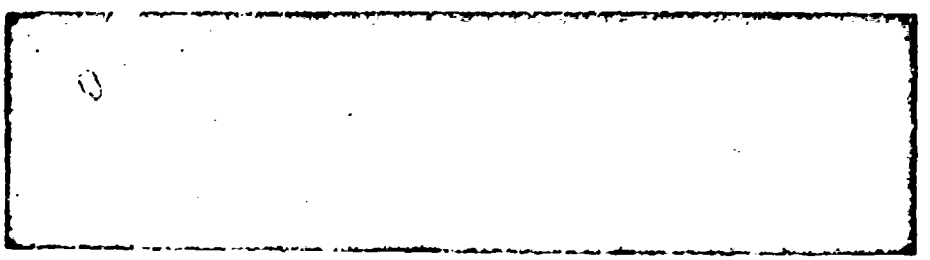


Figure 10. Shroud Antennas on Spearobee

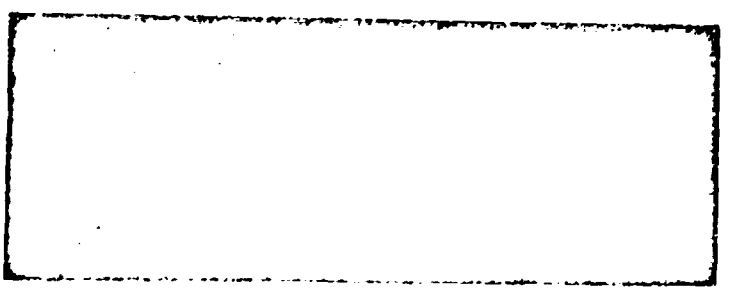


Figure 11. Shroud Antennas on Aerobee-Hi

radiation efficiency at 38 and 76 Mcs has been measured by comparison with a half-wave dipole. At 76 Mc, the ratio of dipole to Shroud radiation is approximately 2 to 1 in voltage, (6 db) which corresponds to 4 db below an isotropic antenna. At 38 Mc, the ratio of dipole to Shroud radiation is approximately 5.5 to 1 in voltage, (15 db) which corresponds to 13 db below an isotropic antenna. Since the maximum transverse projection is only 1.5 inches from the rocket body ($\lambda/200$ at 38 Mc) on a rocket of only six inches in diameter ($\lambda/50$ at 38 Mc), the above efficiencies are far superior to a dipole of similar dimensions.

Continuous signal strength records have been made at the ground receiving sites for 47 of the rockets carrying the Shroud antennas. The average value of received signal is in very good agreement with the calculated value using the above efficiency values.

CONCLUSION

The Shroud antenna described can be conveniently installed on rockets of varied configurations and dimensions and designed for operation at any frequency in the VHF range. Design criteria are simply that the shroud length be of the order of one-quarter wavelength for maximum efficiency. The radiation pattern is independent of rocket length and has been satisfactory with rocket diameters up to one-half wavelength. The basic physical loop configuration lends itself to a rugged, stable mechanical design to withstand the environment of high velocity rockets.

ACKNOWLEDGMENT

The original design concept and subsequent practical development of the Shroud-loop antenna system is a result of the effort of the members of the Radiation Section of the Ballistic Research Laboratories. Noteworthy contributions were made by Mr. Cecil L. Wilson and Mr. Kenneth A. Richer. Mr. Harry Morgan, a Technical Representative of the Philco Corporation, also made many valuable contributions to the design and installation techniques of the Shroud antenna while working at the Ballistic Research Laboratories and at the Fort Churchill IGY Rocket Range.

REFERENCES

1. "DOVAP Antennas for Pre-IGY", by Kenneth A. Richer, BRL Memorandum Report No. 1087, dated July 1957.

2. "Ballistic Instrumentation and Summary of Instrumentation Results for the IGY Rocket Project at Fort Churchill", by R. W. Zamsonats, BRL Report No. 1091, dated January 1960.

3. "Determination of Phase Centers and Amplitude Characteristics of Radiating Structures", by Tetsu Morita, SRI, Technical Report 1, Project 898, dated March 1955.

4. "DOVAP Antenna Recommendations for the International Geophysical Year Program", by C. W. Steele, SRI, Technical Report 4, Project 898, dated 15 January 1957.

5. "Pattern Characteristics of DOVAP Missile Antennas", by C. W. Steele and T. Morita, SRI, Technical Report 5, Project 898, dated August 1957.

6. "Determination of Phase Centers and Amplitude Characteristics of Radiating Systems", by Tetsu Morita, SRI, Final Report, Project 898 - A Summary of Reports Issued on the Program, dated July 1957.

7. "Installation and Adjustment of Shroud-Type DOVAP Antennas on Cajun Rockets", by Spence T. Marks and Harry Morgan, BRL Technical Note No. 1270, dated July 1959.

CONTINUOUS WAVE RANGE SYSTEM

Frank C. Lamm
Philco Western Development Laboratory
Palo Alto, California

INTRODUCTION

The classical means by which range measurements are taken are well known in the field of radar; briefly, a propagated pulse is transmitted to an object and back to its source to measure the time interval.

The space age has brought about a demand for more stringent technological requirements for range and tracking systems. A new system of measurement has been developed to meet the demands of accurately positioning a vehicle in space at ranges greater than 3000 nautical miles. This system is based on the fact that range can readily be computed if the distance can be measured as a function of time.

The range system to be discussed is a continuous wave system with no relation to Doppler-type range devices. The discussion will be concerned with a continuous wave range system designed, fabricated and tested for a maximum range of 3000 nautical miles and an accuracy of 0.05 nautical miles. However, these figures do not represent the maximum limits in range or accuracy that can be obtained by this system.

CW RANGING THEORY

Figure 1 will serve to illustrate the following analysis:

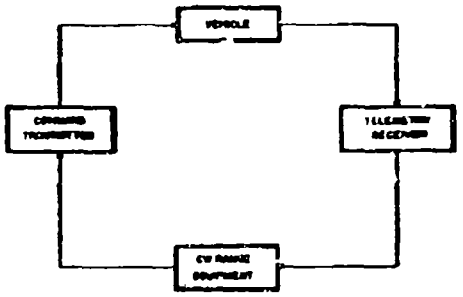


Figure 1

The system is comprised of a ground based transmitter, receiver, and airborne transceiver, all of which are normally necessary for other weapon system functions, voice channel, television and telemetry. The additional equipment needed is the range equipment.

The system is instrumented around the range equipment which measures the round trip transit time of a radar frequency signal from the ground tracking station to the airborne transceiver. This is accomplished by generating four low frequency crystal controlled frequencies in phase coherence and synchronism which modulate the command transmitter. The r-f wave in turn traverses the round trip distance to the vehicle. The received r-f signal is phase-compared to the signal modulation being transmitted, and the resultant information is the slant range. The information is accumulated in binary and decimal form for analytical and visual use.

RANGING SYSTEM ANALYSIS

A detailed description of the ranging system, including block diagrams and a summary of system parameters, is presented in the following discussion.

A practical continuous wave range system must have several sub-carrier ranging tones. The lowest frequency selected must satisfy the requirements for the maximum range to which the equipment must operate. The highest tone must be selected to satisfy the accuracy imposed on the system. The two intermediate frequencies, together with the lowest frequency, are used to resolve the unambiguous range measurement.

RANGE MEASUREMENT EQUIPMENT

A block diagram of the ranging equipment is given in Figure 2.

The range tones are derived from a temperature controlled crystal oscillator. For a range of greater than 3000 nautical miles and an accuracy of 0.05 nautical miles, the stability of the oscillator must be better than 1 parts in 10^6. The intermediate frequencies and low frequency range determining tones are derived from a synchronous frequency divider. The stability of these tones will have the same stability as the crystal oscillator. The ranging tones are sent through phase coherent circuits which filter and phase lock them to the desired angle. The tones are then sent to a precision phase compensator which compensates for the time delay in the range loop. The compensated tones, in phase lock, are modulated on a sub-carrier and fed to the multiplexer and command transmitter.

The ranging tones, before modulation, are sent to an ambiguity resolver which selects the proper cycle of the highest tone to initiate the range measurement.

The modulated r-f signal is transmitted to the vehicle receiver and back to the ground station. After traversing their round trip, the signal is demodulated and filtered to extract the ranging tones. The delayed tones are then sent to a similar ambiguity resolver which selects, by comparing the phase of the intermediate and low frequency tone, the proper high frequency period to terminate the range interval. The interval is converted to a binary word which represents

the slant range and is then stored for computer use. The range word is also converted to decimal form for visual indication and analogue form for presentation on an oscilloscope.

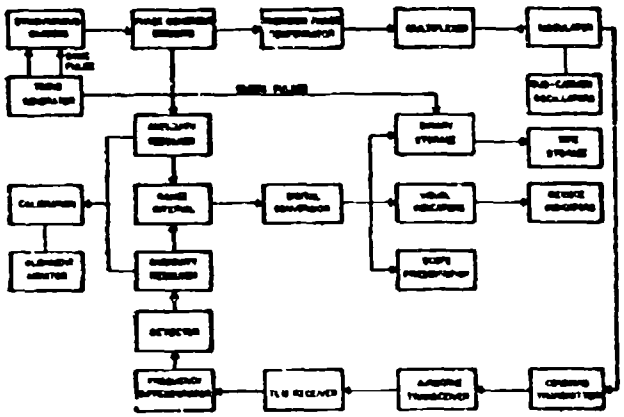


Figure 2

PARAMETER SELECTION

High Frequency Tone. The selection of the highest frequency is based on the following requirements: accuracy requirements of the system, space allotted on the telemetering base-band, and ability to measure time delay. An accuracy of 0.05 nautical miles can be achieved by selection of a frequency between 25 and 30 K cps. The phase stability of the system can then be held to within 4 degrees without complex circuit problems. The required telemetry bandwidth will remain well within 200 K cps.

Intermediate Frequency Tones. The selection of the intermediate frequencies are a compromise between proper instrumentation of the system, and stringent phase requirements which would result in selection of tones separated from the highest tone by an excessive interval. A ratio of 10:1 but not greater than 15:1 is a practical compromise.

Low Frequency Tone. The selection of the low frequency tone is governed by the maximum range desired from the equipment. The wave length selected shall be greater than the maximum range to be measured, as shown below.

$$\lambda = 2 L_m + \emptyset$$

where λ = wave length of lowest frequency

L_m = maximum station to vehicle distance

$\emptyset$ = the time required for the range equipment to compute and readout the range word

TIMING GENERATOR

The ranging tones must satisfy the following requirement in order to obtain a high accuracy system:

1) Stability of 2 parts in 10^6 for ranges of 3000 to 6000 nautical miles.

2) Stabilities of 1 part in 10^6 for ranges of 6000 nautical miles to 9000 nautical miles.

3) Stabilities of 9 parts in 10^7 for ranges of 9000 nautical miles to 12000 nautical miles.

4) Phase coherence of frequencies.

5) Synchronism of frequencies.

The stability of the tones are dependent on the crystal controlled oscillator. By the use of a synchronous frequency divider to generate the intermediate and low frequency tones, identical stability of all the tones is assured. Phase coherence of the tones are also satisfied by this method.

Synchronization of the tones will place the range information on a real-time axis. By initiating each range measurement with a sync pulse which is compared to ground time and adding the resultant range interval to this, the exact time of acquisition can be determined. The range word supplied to the computer will then be on a real-time axis.

ANALYSIS OF PHASE ERRORS

The phase errors which contribute to the overall accuracy of the system are categorized below:

1) Signal-to-noise ratio

2) Doppler frequency shift

3) Calibration error

4) Changes in propagation velocity

5) Random phase shifts

A thorough and complete analysis of the above areas will not be attempted in this paper; however, the pertinent points will be discussed.

SIGNAL-TO-NOISE PHASE ERROR

The phase error ($\Delta\phi_n$) due to a finite signal-to-noise ratio can be explained by the following expression:

$$\Delta\phi_n = \frac{1}{\sqrt{2\,\frac{S}{N}}}$$

where $\Delta\phi_n$ = rms phase error in radians

S/N = numeric rms power ratio

DOPPLER FREQUENCY SHIFT ERROR

The expression for doppler phase error for a circular orbit is as follows:

$$\Delta\phi = \frac{2\pi\, lm}{C^2}\, V\cos\psi$$

where $\frac{\omega}{2}$ = frequency

lm = range

$V\cos\psi$ = radial velocity

C = propagation velocity

CHANGES IN PROPAGATION VELOCITY

The path of propagation is increased and the velocity of propagation decreased in tracking an orbiting vehicle. The apparent increase in path length is a function of the carrier frequency. The 3000 nautical mile system, discussed when operating with a carrier frequency of 1.5 Kmc to 2.5 Kmc, will realize an inherent path change of approximately 10 to 25 meters.

CALIBRATION PHASE ERRORS

The phase error contributed to calibration of the range equipment is directly related to the resolution of the equipment and the accuracy of the precision phase compensators. Maximum ranges of greater than 3000 nautical miles are pre-calibrated to within 10 meters. Where applicable, the use of high resolution techniques will reduce this figure to 1.5 meters.

CALIBRATION

The calibration of the range loop is accomplished by use of a ground-based vehicle simulator. The ranging tones are operating on and fed to the command transmitter for transmission to the vehicle simulator. After completing the loop through the telemetry receiver, the system time delay is read in the range equipment. The precision phase compensators are then adjusted to indicate zero range in digital form. The pre-calibrated time delay through the coaxial lines to the vehicle simulator is then added to the zero indication to give true zero range.

The calibration mode is relay controlled and is completed in several minutes. The calibration mode is also used if any breakdown is suspected before a vehicle data pass. This enables thorough and complete checkout of equipment before its use.

RESOLUTION

The accuracy of the range measurement under ideal conditions is approximately 20 meters. This is determined by precision in which the range interval can be resolved assuming the above contributing errors are negligible. With range measurements greater than 200 nautical miles, 20 meter resolution is more than adequate for accuracies of 0.05 nautical miles. However, in special applications where shorter ranges are required, a resolution of 1.5 meters can be obtained. This is accomplished by specialized phase-lock digital techniques incorporated in the equipment for high resolution. The additional circuitry required to obtain this increased resolution is small in comparison to the present range equipment. The block diagram in Figure 3 serves to illustrate the technique incorporated for a high resolution system.

NOISE BANDWIDTH

The use of specially designed narrow-band filters for extracting the ranging tones reduces the noise bandwidth of the system to levels approaching that of a micro lock receiver. Narrow noise bandwidths make possible the extension of range in the existing system to greater than 15,000 nautical miles.

EXTENDED RANGE

The above described system is capable of measuring range in excess of 15,000 nautical miles. However, with the advent of moon orbits, the requirements for accurate ultra-range tracking equipment is immediate. The feasibility of resolving ranges as great as 240,000 miles is made possible with a continuous wave range device.

Extending the limits of the measurement by transmission of tones below several cycles is not a practical solution; however, experiments have been conducted to extend the range of the 15,000 nautical mile system. One method that may prove practical is the placement of range markers on the ranging tones. The markers will serve to select the proper quadrant in which the vehicle is traveling which in this case is represented by a sector of range 15,000 nautical miles in length. The addition of range markers will not increase the bandwidth requirements of the system. The accuracy of the system will, however, be relaxed to compensate for the greater effects due to the previously mentioned errors.

ACKNOWLEDGMENT

I would like to acknowledge the contributions of the following people:

L. W. Cartwright
J. Stephenson
D. Smith
J. Grimaldi

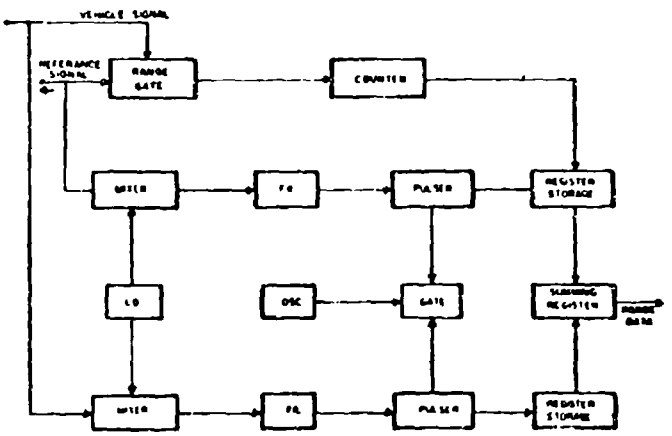


Figure. 3

AN EXPERIMENTAL STUDY OF MONOPULSE RADAR FOR GROUND CLUTTER DISCRIMINATION

By: Shurman Y. Chang and Vincent Stabilito, Fire Control Div., Frankford Arsenal

I. INTRODUCTION

The objective of this study is to experimentally investigate the capability of a phase-comparison monopulse radar in distinguishing between ground target and clutter for application to surface-to-surface detection and ranging of military targets in fire control systems.

The main problem of surface-to-surface radar for detection and ranging of military targets is the effect of clutter caused by natural objects on the ground such as trees and bushes. The experience of testing scan modulation type radars such as T-39, T-44 and T-47 in the past few years has unmistakably shown that the ability of this type radar to discriminate between a target and its clutter environment is insufficient, thus making it almost impossible to determine the range of a target in clutter environment under practical field conditions. Although a number of doppler-type ground radars such as PPS-4, TPS-21 and TPS-25 have been developed to detect moving targets, no radar is available today which can distinguish between stationary targets and ground clutter.

The monopulse phase-comparison technique represents one approach to minimize the effects of vegetation-type clutter. In the spring of 1958, Frankford Arsenal initiated an experimental program to investigate the capability of phase-comparison monopulse radar for possible clutter discrimination. A breadboard radar was built in the laboratory and later evaluated at a selected site at Ft. Dix, New Jersey. The description of the breadboard and some early data taken was presented in a paper at the 1959 Army Science Conference held at the Military Academy, West Point, New York. Considerably more data were taken during May through mid-July 1959, after the original experimental equipment had been somewhat modified. The results of the study have indicated not only an improvement in clutter discrimination in comparison with T-47, a conical scan range-only radar, but also a definite pointing characteristic of stronger target preference.

II. TECHNICAL DISCUSSION

A. Description of System

The breadboard equipment used for the experimental study can be best described with reference to the block diagram and photographs as shown in figures 1 and 2. The system characteristics are given in figure 3. The system consists of dual paraboloids with feed centers displaced by 12 3/8" (36λ) in the azimuth plane for phase comparison monopulse operation and dual receivers to yield sum and difference video signals. The transmitting horn was used mainly for illumination of target during the early part of the study and for facilitating antenna pattern measurements. The data to be presented were actually taken with duplexers in the system.

The target echoes from the receiving antennas are fed into a magic-tee to produce sum and difference radio frequency signals. Since the feed centers of the dual parabaloids are separated by a fixed amount in the azimuth plane, the signal voltages at the feed centers have the same magnitude but a difference in phase proportional to θ of very small values, where θ is the angle between the direction of target and the center line of the antenna system. It can be shown that the magnitude of sum signal voltage from the output of magic-tee follows an even function of θ and that of the difference signal voltage follows an odd function of θ. When the target is on the center line of the antenna system, the sum signal voltage reaches a maximum while the difference signal voltage becomes a null. By means of an electronic switch, both the sum and the difference video signals can be displayed simultaneously on an oscilloscope.

Box-car circuits were used in the system to convert the sum and difference video into box-car so that the sum-channel box-cars and the difference between the sum and difference channel box-cars, obtained from a summing circuit, could be displayed and photographed.

In addition to displaced feed-centers, the two antennas were also adjusted in the azimuth plane to yield a desired operating characteristic such that:

(1) There was no squint in the elevation plane, i.e., parallel beam axes in elevation plane.

(2) There was a squint in the azimuth plane that means beam axes were not parallel but divergent in this plane. The amount of squint was adjusted to affect maximum reduction in secondary lobes in the difference signal pattern in order to provide one null or a minimum corresponding to the sum signal maximum when the antenna beam was scanned in azimuth across an isolated target.

The resulting individual antenna patterns and the sum and the difference signal patterns for one-way transmission are illustrated in the photographs shown in figures 4 and 5. The sum and difference signal patterns with duplexer in the system is shown in figure 6. The individual patterns were

taken in two separate scans with the two antennas alternately blocked out.

It should be noted in the above figures that the peak of the sum signal and the null of the difference signal appears in line with the cross-over points of the two individual patterns. As a result of squinting the antennas, there was a separation of 22 mils between the two individual patterns.

B. Presentation of Experimental Results

During a three months period, partly in late winter and partly in early summer of 1959, experimental data were taken at Ft. Dix, New Jersey. The test area at Ft. Dix, selected for the experimental investigation, was relatively flat terrain. Bordering on one end of the area at a distance of about 900 yards from the radar site were pine trees, bushes, weeds and grass as shown in figures 7 and 8. At closer range from the radar site up to about 800 yards, the terrain had sparsely scattered bushes and weeds. A 1 1/2 ton Army truck was employed to house the radar equipment, corner reflectors were used as targets as well as for setting up reference levels. The only vehicle-type target employed in this experiment was a commercial station wagon. For reference and comparison, a T-47 conical scan radar was utilized at the same site.

For the convenience of classification, the data are divided into two categories, namely, (1) data indicating clutter-discriminating characteristics and (2) data indicating target-pointing characteristics. They are shown separately as follows:

(1) Data Indicating Clutter Discriminating Characteristics

Box-car signals for vegetation type clutter both from the output of summing circuit and from that of sum-channel box-car circuit are shown in figure 9. It appears that the magnitude of the box-cars from the sum-channel remained positive and practically constant while that from the summing circuit showed both positive and negative values and the averaged value is rather small but positive. This phenomena probably results from the change in phase relationship between the signals at the antenna sensing points on account of the fluctuations of the random scatterers which make up the clutter. If a corner reflector were employed as a target, the two box-car signals corresponding to the above figures would be practically identical because the null in the difference signal for on-axis target would not cause any cancellation to affect the sum-channel box-cars.

Another set of results indicating clutter discriminating characteristics is the simultaneous display of the sum and difference video signals as shown in figures 10, 11 and 12. A-displays by using T-47 reference radar are shown in the upper parts of these figures. The clutter area in which the corner reflector was located has been shown previously in figures 7 and 8. The white rectangular object visible in these figures marked the location of the reflector. A station wagon in clutter is shown in figure 12 together with A-displays for both reference and monopulse radars. The peak following the target signal was due to the ridge of ground running diagonally across the rear of the vehicle. A marked improvement in clutter discrimination by monopulse radar, as compared with the reference radar, should be noted in the above three figures.

Numerically, the signal to clutter contrast can be expressed in terms of a normalized clutter area as defined and calculated in the Appendix. In figure 11, where a 10-M^2 corner reflector was used as a target, the normalized clutter area was found to be 0.00037 M^2 per M^2 of terrain area while the value for the reference radar was 0.0744 M^2 per M^2 of terrain area approximately 23db larger. The station wagon used in figure 12 can be regarded as not significantly different from a military vehicle when used as a radar target. To detect and locate the vehicle as a target is practically impossible in the case of a conical scan radar on account of the severe modulation effects caused by the ground clutter. In contrast, the monopulse radar unmistakably detected the existence of a hard target in clutter by virtue of the maximum sum and null difference characteristics so distinctively exhibited in the photograph.

(2) Data Indicating Target Pointing Characteristics

The pointing characteristics of the experimental monopulse radar are illustrated by a series of photographs of sum and difference signal amplitude vs. azimuth angle for two point-source targets of different sizes set up at a given range but having various angular separations. Figures 13, 14 and 15 show this series of photographs. Pointing simply refers to the point on the amplitude vs. azimuth display where the null of difference signal coincides with the peak of the sum signal.

A corner reflector was first placed approximately at a range of five hundred yards from the radar. The latter was then moved in azimuth until a null was indicated in the difference signal. The null was used as zero reference which

permitted collimation of radar and optical axis. The antenna position reference was finally adjusted for centering on the oscilloscope. The vertical center line shown in the above photograph pointed the position of this target, the larger of the two reflectors. The second reflector, about 3db smaller in radar area, was located approximately at the same range but at various angular separations with respect to the first, starting from 2.5 milliradians and increasing up to 55 milliradians.

For the smallest separation 2.5 milliradians shown in figure 13, the radar pointed at the larger of the two targets within the beamwidth. The presence of the smaller target was hardly recognized. At 5 milliradians separation, only the larger target was pointed at. The appearance of secondary lobes in the difference signal was also noted as a result of the presence of the smaller target. The increasing area of the left half of the difference pattern, was the evidence of the presence of the second reflector while the pointing at the larger target remained unchanged. A second null began to be noted in figure 13 for 25 milliradians separation but there was no corresponding peak in the sum pattern. Scintillations were evident due to wind effects which caused relative motions of the targets. The larger target was still pointed at accurately. The second null appeared more plainly in figure 14 for 30 milliradians separation. Same as the case for 25 mils separation there was a 5-milliradian pointing error for the second null. The pointing at the main target was unchanged. Two targets were seen at separate nulls without any pointing error when the separation was increased to 35 milliradians corresponding to a single antenna beamwidth. The same phenomena was noted in figure 14. Correct pointing at the larger target was maintained for 45 milliradians. However, a third null now made its appearance. A similar phenomena was shown in figure 15. In both cases there was an apparent pointing error of 1.5 milliradians at the smaller target probably attributable to the non-linear antenna position reference employed.

In order to verify the experimental results on pointing characteristics, mathematical expressions for the magnitude of the sum and the difference signals, Es and Ed, have been derived as a function of relative sizes of two targets, angular separation between them, squint angle of the antennas, difference in ranges of the targets and direction of the stronger target with respect to the center line of the antenna system. (See Appendix II for the expressions.) Three sets of curves are shown in figures 16, 17 and 18 for different angular separations between two targets, for a fixed squint angle and a fixed ratio of target sizes. The average curves on the plots were obtained by taking the average of the ordinates of different curves. These curves indicate practically no shifting of null-point for wide angular separations but a shifting in the magnitude of 1-mil in the case of 5-mil separation. A distinct pointing characteristic of strong target preference is demonstrated by these analytical results closely resembling the experimental results as discussed above. The average curve for the sum signal has been plotted separately (now shown in the Appendix) for 15-mils separation. The maximum on the curve coincides with the minimum of the corresponding average curve for difference signal.

During the later phase of the study an interesting observation was made of A-scope presentation of a large complex structural target viewed at different angles from the radar and at a range of about 1500 yards. The target is shown in the photograph in figure 19. Figure 20 shows a plan view of the target of interest. Superimposed on the plan view are four dashed lines representing the directions of the radar designated at -3-mil, 0-mil, 5-mil and 10-mil.

Four corresponding pairs of traces, each presenting A-displays of sum and difference video signals, are shown in figure 21. The solid heavy lines in figure 20 represent simulated intensity modulations based on all ratios, exceeding four, of sum over difference signal voltages. The lines are plotted according to the four angular positions of the target viewed by the radar, but the superposition is made arbitrarily in range. The small pulse at the left end of the trace in the A-presentations was obtained from a telephone pole in the foreground of the target now shown in the photographs. It is short enough to permit indexing of the sum and difference traces in time. The voltage readings were taken at an interval of every one-third of a pulse length. Any ratio of the sum and difference voltages over four registered a solid spot of 9' in range along the dash lines in figure 36. During the test the gain settings of the sum and difference channel amplifiers were adjusted as close to be the same as possible. Attention should be drawn to the ability of the radar to resolve a large complex target into various separate bits of information by means of scanning and ratio gating techniques. In case of conventional radar, the same target would be presented on an A-scope in the form of a single long broad pulse without revealing its nature and complexity.

III. SUMMARY AND CONCLUSIONS

Based on the results of the experimental study and the evaluation thereof, the following summarized conclusions can be drawn:

a. The sum-channel box-cars of vegetation type clutter have positive and practically constant amplitude, but the box-cars from the summing circuit for the difference between sum channel and

difference channel box-cars have both positive, negative and zero values on account of the random phase relationships of the two radio frequency signals at the sensing points. For a point-source target, this same box-car would be positive and practically constant in amplitude much resembling the clutter.

b. There exists a distinct clutter discriminating characteristic in the monopulse system. The coincidence of the maximum of the sum video and the null or the minimum of the difference video establishes a clear cut condition for the on-axis target even when it is set in heavy clutter. Numerically, the monopulse system is capable of a 23db improvement in clutter discrimination over a conventional or conical scan type radar of comparable characteristics.

c. The experimental system has a pointing characteristic of strong target preference. For two targets of different sizes, within the beamwidth, the null or the minimum of the difference signal points accurately at the stronger target for any angular separation between them from a few mils to more than a beamwidth where a second null appears to indicate the location of the smaller target. This characteristic has been supported and verified by analytical results.

APPENDIX I

1. The normalized clutter area is defined as the ratio of clutter area, determined by the relative clutter and target signal powers and target area, to the physical area of the radar cell defined by the product of pulse duration and beamwidth at a fixed range. Small grazing angle is implicit in the above definition.

The signal to clutter voltage ratio is given as

$$E_\Sigma - E_\Delta/E_\Delta$$

where E_Σ and E_Δ are respectively the sum and difference video signals from the lower portion of the photograph in figure 1

Numerically the above ratio is 9 and the corresponding power ratio is 81 which indicates the relative target signal and clutter power. For the corner reflector used, the target area is $10m^2$. The clutter area then becomes $10/81 = 0.123m^2$. The physical area illuminated by the radar, i.e. the radar cell, is found to be $329.6m^2$ at the range of 871.5 meters, pulse length of 8.75 meter and beamwidth of 2.58^0.

Therefore, the normalized clutter area as defined above can be calculated as follows:

$$\frac{0.123m^2}{329.0m^2} = 0.00037m^2/m^2$$

2. In a similar manner the normaliz clutter area, when the control radar is used, can be calculated with reference to the photographs in figure 10. Its value is found to be $0.0743m^2/m^2$.

The ratio of the normalized clutter area for the monopulse and control radars, using the latter as reference, is found to be

$$\frac{0.0743m^2/m^2}{0.00037m^2/m^2} = 200 \text{ approximately}$$

or 23db difference.

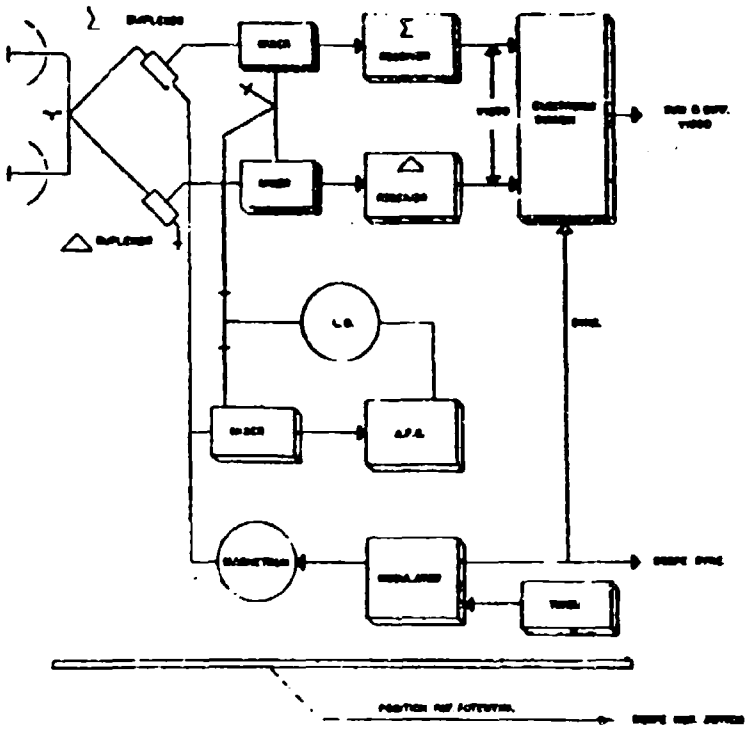
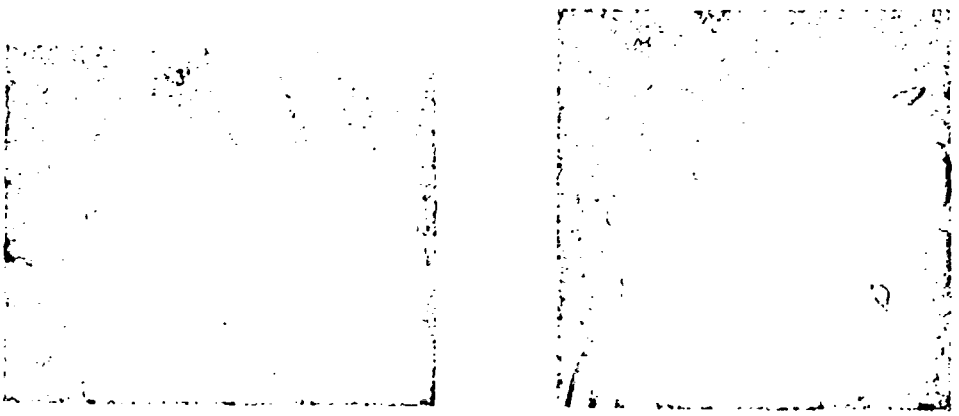


FIGURE 1. MONOPULSE BREADBOARD BLOCK DIAGRAM TRANSMITTER AND RECEIVERS



SYSTEM PARAMETERS AND CHARACTERISTICS

TRANSMITTER FREQUENCY	36,000 MC/S
PEAK TRANSMITTED POWER	35 KW
PULSE DURATION	0.08 MICROSECONDS
PULSE REPETITION FREQUENCY	4,000 P/S
TRANSMITTING HORN	
GAIN	23 DB
BEAMWIDTH	10° IN AZIMUTH 10° IN ELEVATION
RECEIVING ANTENNAE (DUAL PARABOLOIDS)	26" x 26"
GAIN	44 DB
BEAMWIDTH	3° IN AZIMUTH 3° IN ELEVATION
I.F. FREQUENCY	60 MC/S
I.F. BEAMWIDTH	20 MC/S
OVERALL NOISE FIGURE / CHANNEL	10 DB

FIGURE 3. SYSTEM PARAMETERS AND CHARACTERISTICS

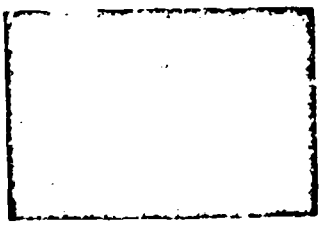


FIGURE 4. INDIVIDUAL ANTENNA PATTERNS NO DUPLEXER, 125M² CORNER REFLECTOR, LOWER PATTERN FOR HIGHER GAIN

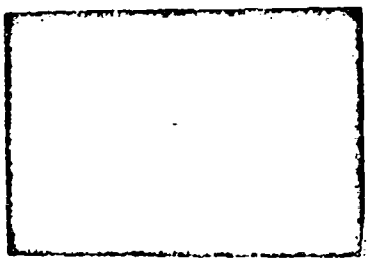


FIGURE 6. SUM AND DIFFERENCE SIGNAL PATTERNS WITH DUPLEXER 125M² REFLECTOR 500M RANGE 35 MILS PER DIVISION



FIGURE 5. SUM AND DIFFERENCE SIGNAL PATTERNS - NO DUPLEXER 125M² CORNER REFLECTOR 500M RANGE, 35 MILS PER DIVISION



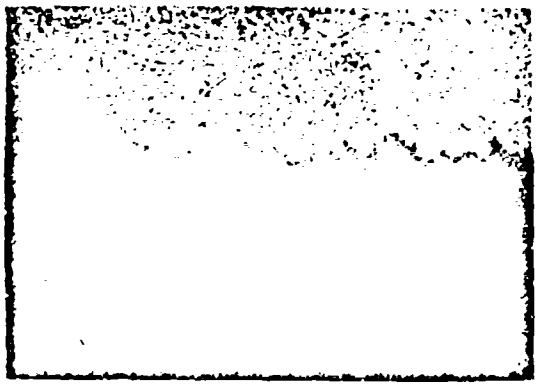
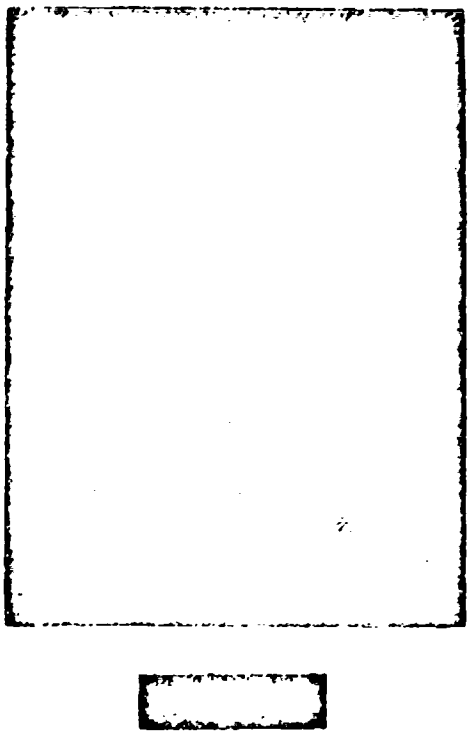


FIGURE 8. CLUTTER AREA: HEAD-ON CLOSE-UP VIEW LATERAL DISTANCE APPROX. 24 YDS. EQUIVALENT ANGLE 25 MILLI-RADIANS



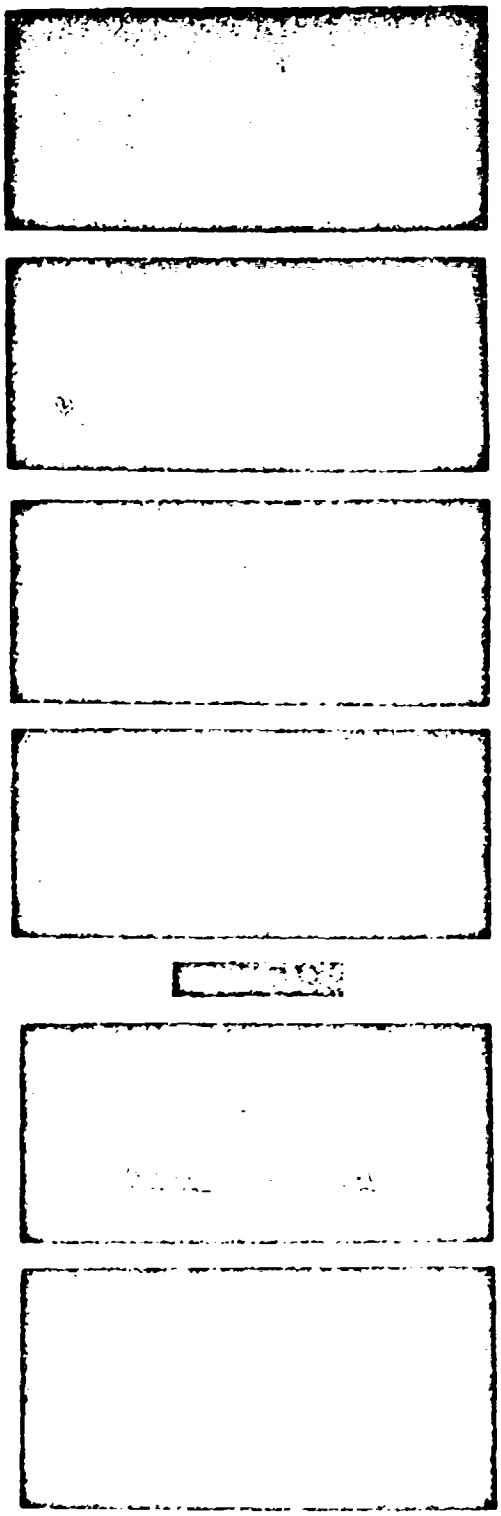
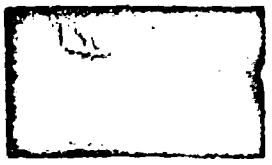
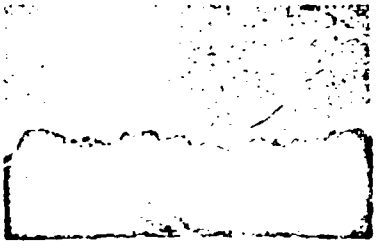


FIGURE 11.



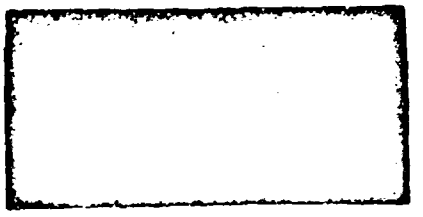
STATION WAGON A-DISPLAY
T47 RADAR - 750M RANGE

STATION WAGON A-DISPLAY
MONOPULSE RADAR - 750M RANGE



STATION WAGON IN CLUTTER
750M RANGE

FIGURE 12.



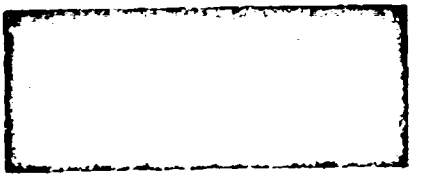
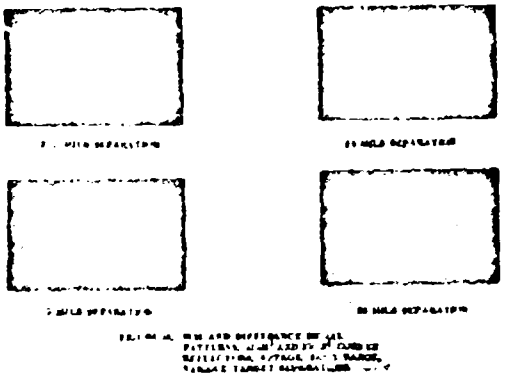
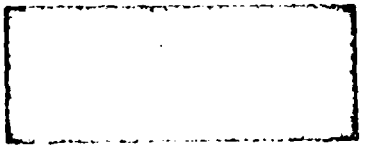
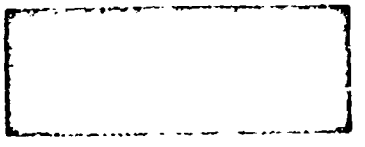


FIGURE 13. SUM AND DIFFERENCE SIGNAL PATTERNS, 125M² AND 250M² CORNER REFLECTORS, APPROX. SAME RANGE, VARIOUS TARGET SEPARATIONS

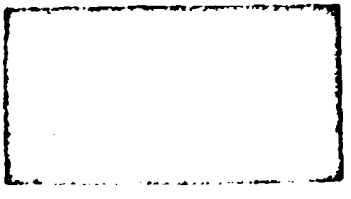




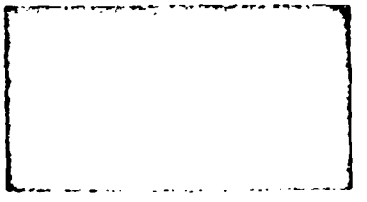
30 MILS SEPARATION



40 MILS SEPARATION



35 MILS SEPARATION



45 MILS SEPARATION

FIGURE 14. SUM AND DIFFERENCE SIGNAL PATTERNS, 125M² AND 250M² CORNER REFLECTORS, APPROX. SAME RANGE, VARIOUS TARGET SEPARATIONS

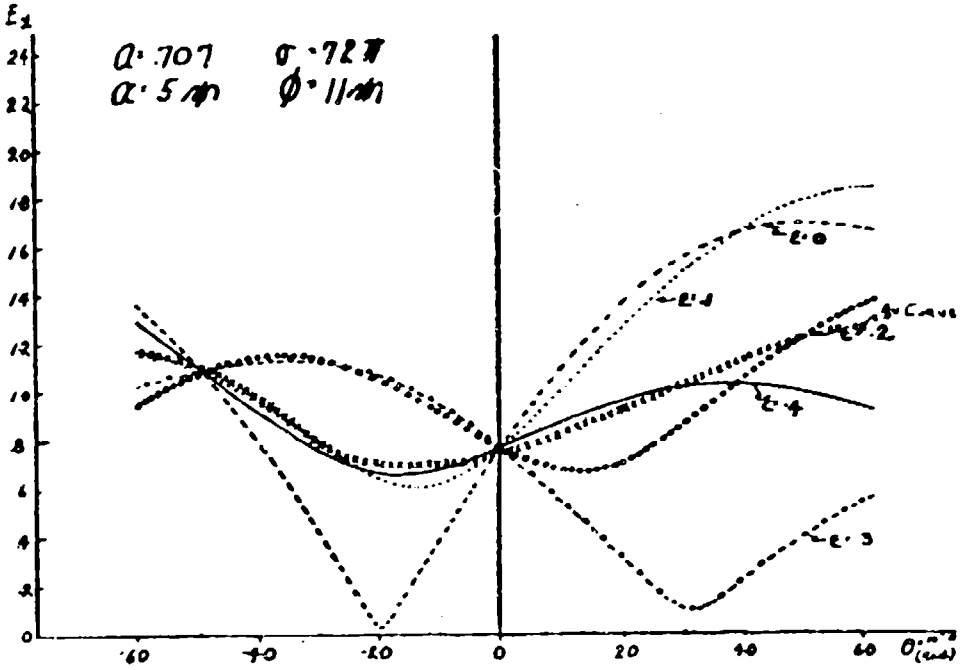


FIGURE 16. MAGNITUDE OF DIFFERENCE CHANNEL VIDEO AGAINST DIRECTION OF LARGER TARGET

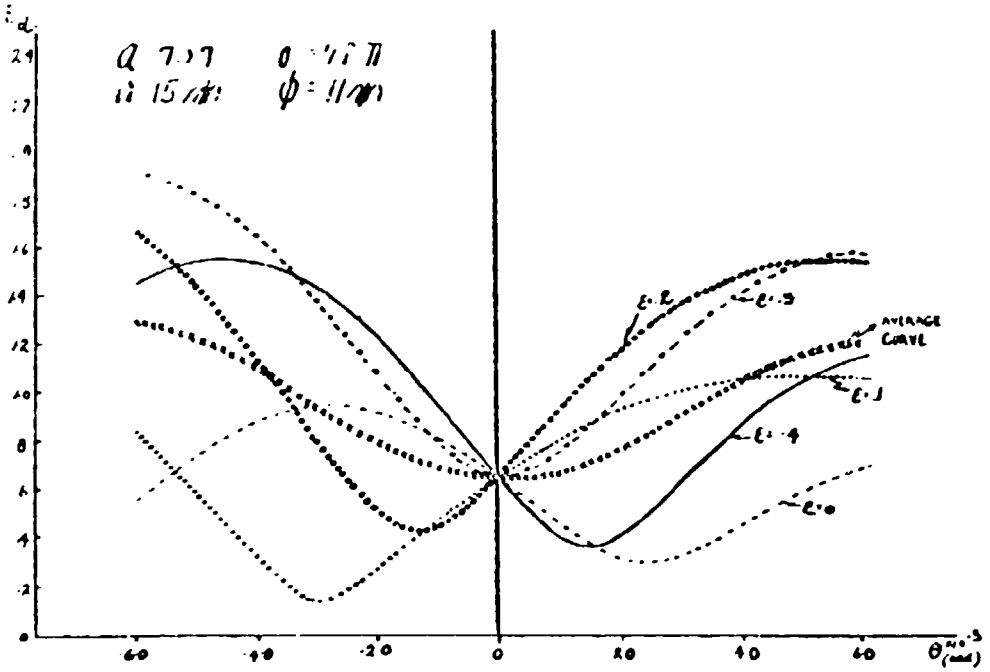


FIGURE 17. MAGNITUDE OF DIFFERENCE CHANNEL VIDEO AGAINST DIRECTION OF LARGER TARGET

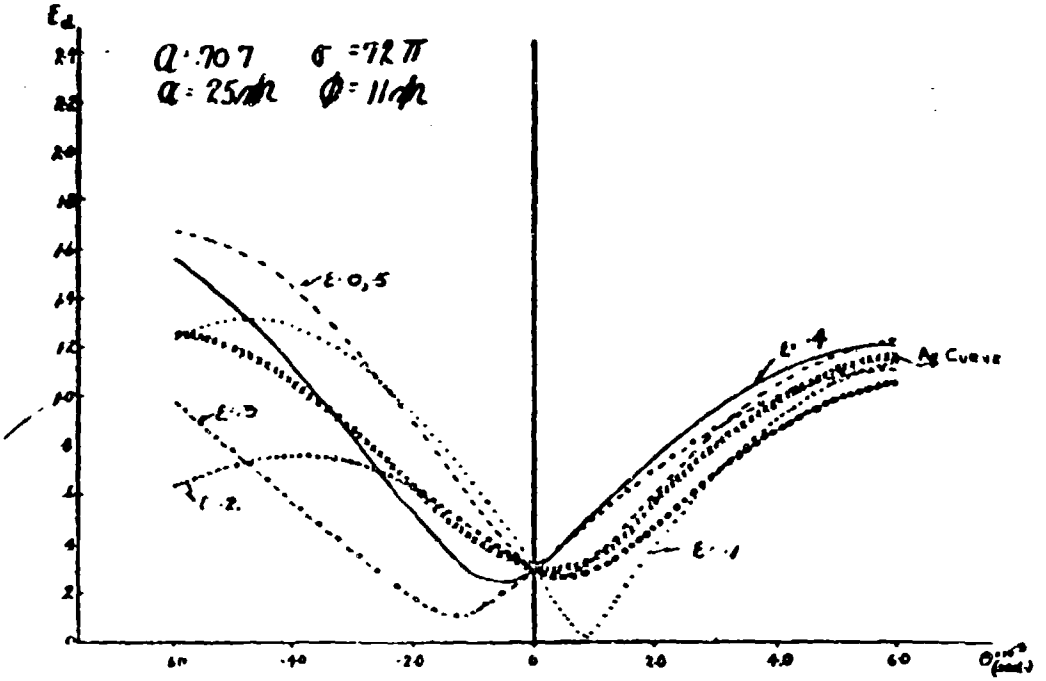


FIGURE 18. MAGNITUDE OF DIFFERENCE CHANNEL VIDEO AGAINST DEFLECTION OF LARGER TARGET

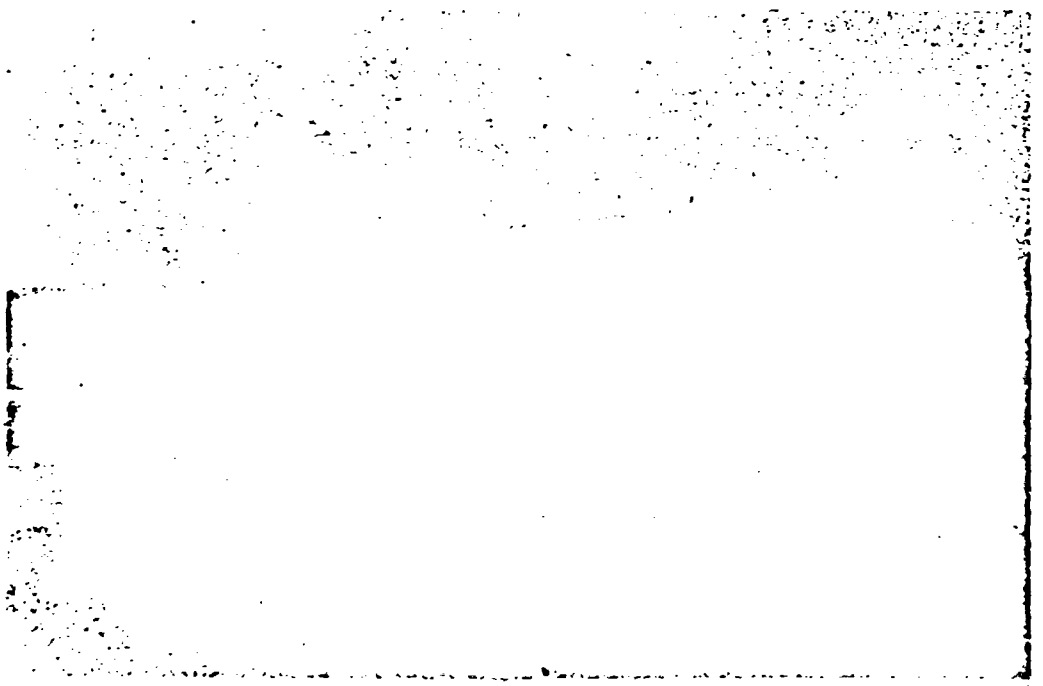


FIGURE 19. A STRUCTURAL TYPE OF TARGET

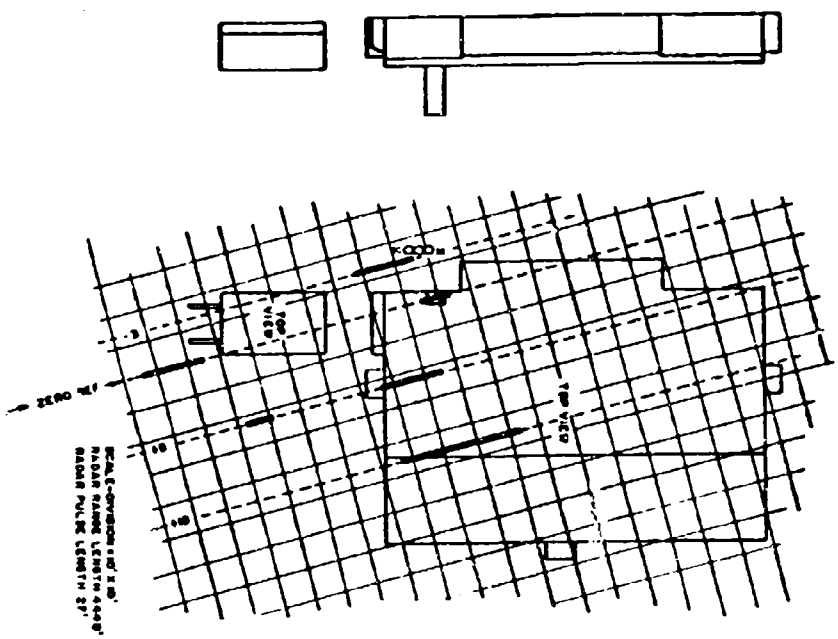
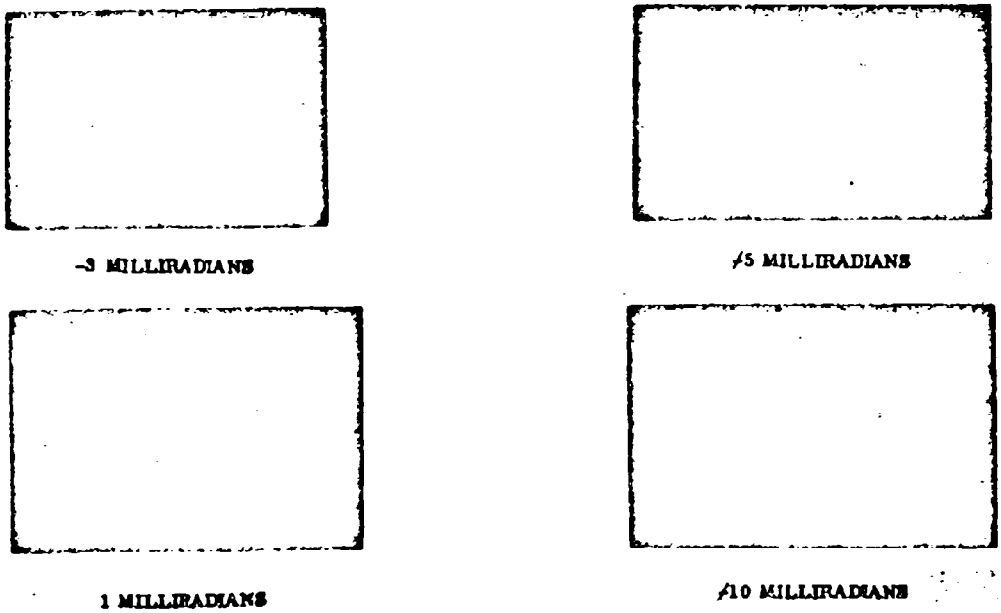


FIGURE 20. PLAN VIEW OF A STRUCTURAL TARGET



-3 MILLIRADIANS

/5 MILLIRADIANS

1 MILLIRADIANS

/10 MILLIRADIANS

FIGURE 21. AMPLITUDE AGAINST RANGE, POINTING DIRECTIONS AS INDICATED, UPPER TRACES - SUM VIDEO, LOWER TRACES - DIFFERENCE VIDEO

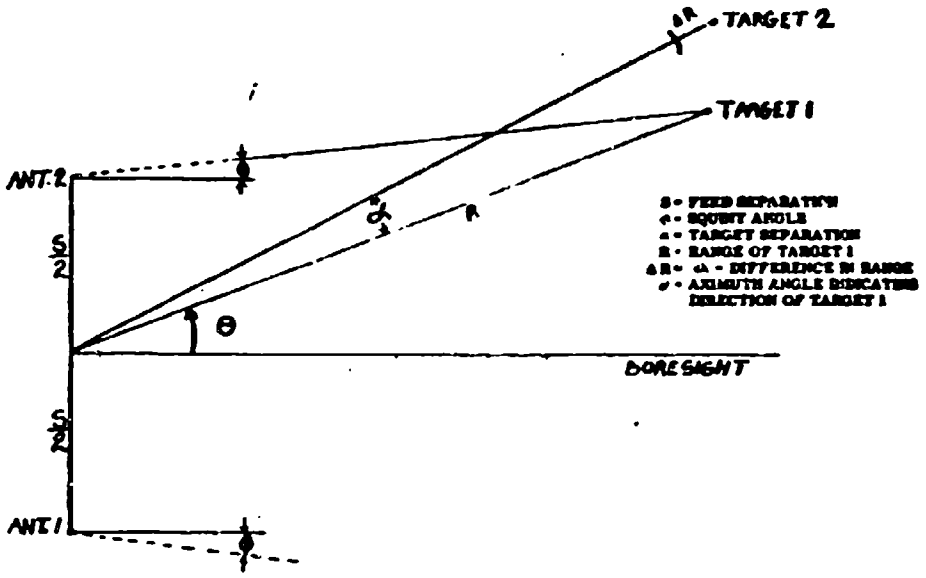


FIGURE 22. GEOMETRY OF A MONOPULSE SYSTEM TWO POINT SOURCE TARGETS

2

$$\phi = 0$$

$$E_d = |e_d| = \sqrt{\left\{\sin(\sigma \sin\theta) + a \sin[\sigma \sin(\theta+\alpha)] \cos 4\pi\varepsilon\right\}^2 + \left\{a \sin[\sigma \sin(\theta+\alpha)] \sin 4\pi\varepsilon\right\}^2}$$

$$\sigma = \frac{(2\pi)}{\lambda} = 72\pi \qquad S = 36\lambda$$

$$\Delta R = \varepsilon\lambda \qquad 0 \leq \varepsilon \leq .5$$

$$a = \sqrt{\frac{A_2}{A_1}} \qquad A_1 = 1$$

FIGURE 23. EXPRESSION FOR THE MAGNITUDE OF DIFFERENCE CHANNEL SIGNAL - NO SQUINT

3

$$E_d = |e_d| = \sqrt{X^2 + Y^2}$$

$$X = \left[\{G(\theta-\phi) - G(\theta-\phi)\} \cos(\sigma \sin\theta) + a \cos 4\pi\varepsilon \{G(\theta-\alpha-\phi) - G(\theta-\alpha-\phi)\} \cos\{\sigma \sin(\theta-\alpha)\} \right.$$
$$\left. - a \sin 4\pi\varepsilon \{G(\theta-\alpha-\phi) + G(\theta-\alpha-\phi)\} \sin\{\sigma \sin(\theta-\alpha)\} \right]$$

$$- Y = \left[\{G(\theta-\phi) + G(\theta-\phi)\} \sin(\sigma \sin\theta) + a \cos 4\pi\varepsilon \{G(\theta-\alpha-\phi) + G(\theta-\alpha-\phi)\} \sin\{\sigma \sin(\theta-\alpha)\} \right.$$
$$\left. + a \sin 4\pi\varepsilon \{G(\theta-\alpha-\phi) - G(\theta-\alpha-\phi)\} \cos\{\sigma \sin(\theta-\alpha)\} \right]$$

$$G(u) = \left[\frac{\sin u}{u} \right]^2 \qquad u = \frac{\pi D}{\lambda} \sin\theta$$

FIGURE 34. EXPRESSION FOR MAGNITUDE OF DIFFERENCE CHANNEL SIGNAL SQUINT ANGLE = 11 MILS

$$E_s = |e_s| = \sqrt{X^2 + Y^2}$$

$$X = \left[\{G(\theta-\phi) + G(\theta-\phi)\} \cos(\sigma \sin\theta) + a \cos 4\pi\varepsilon \{G(\theta-\alpha-\phi) + G(\theta-\alpha-\phi)\} \cos\{\sigma \sin(\theta-\alpha)\} \right.$$
$$\left. a \sin 4\pi\varepsilon \{G(\theta-\alpha-\phi) - G(\theta-\alpha-\phi)\} \sin\{\sigma \sin(\theta-\alpha)\} \right]$$

$$- Y = \left[\{G(\theta-\phi) - G(\theta-\phi)\} \sin(\sigma \sin\theta) + a \cos 4\pi\varepsilon \{G(\theta-\alpha-\phi) - G(\theta-\alpha-\phi)\} \sin\{\sigma \sin(\theta-\alpha)\} \right.$$
$$\left. + a \sin 4\pi\varepsilon \{G(\theta-\alpha-\phi) + G(\theta-\alpha-\phi)\} \cos\{\sigma \sin(\theta-\alpha)\} \right]$$

FIGURE 35. EXPRESSION FOR MAGNITUDE OF SUM CHANNEL SIGNAL SQUINT ANGLE = 11 MILS

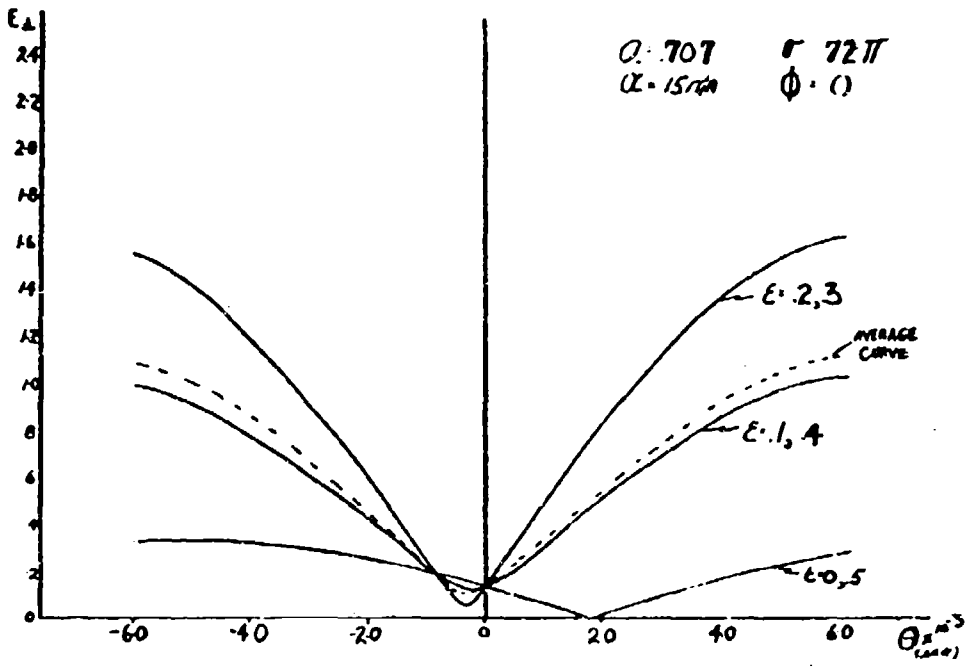


FIGURE 38. DIFFERENCE CHANNEL SIGNAL AGAINST
θ NO SQUINT 18 MILS TARGET SEPARATION

INSTRUMENTATION ERROR ANALYSIS OF THE AMR MISSILE TRACKING SYSTEMS

By: Mr. B. U. Glass, RCA Missile Test Project, Patrick AFB

ABSTRACT

Tracking systems under development for the Atlantic Missile Range are designed to produce absolute accuracies ranging from 1 part in 100,000 for the Mk II AZUSA (soon to be operational) to 1 part in 1,000,000 for the AMR High Precision tracking system (to be operational in approximately three years).

The calibration and evaluation of these systems requires specially designed aircraft and missile tests. The design of these tests and the factors affecting their design are presented. Special emphasis is given to the use of ballistic cameras as the standard for accomplishing the calibration.

Finally, a brief description is given of the mathematical techniques utilized in the analysis of the test data for determining random and systematic errors of the systems.

INTRODUCTION

It is a disturbing fact that conflicting trajectories are obtained from various tracking systems utilized in a given missile flight test. In the early phases of missile development such inconsistencies are generally tolerated by missile contractors because other considerations are of more importance. Indeed, in the early developmental phases relatively little use is actually made of external data, telemetered data being all-important. However, when development has reached more advanced stages, external data are called upon for the derivation of refined propulsion and guidance parameters. In particular, the evaluation of guidance parameters requires external data of extremely high absolute accuracy. Moreover, as guidance systems become increasingly refined, the state of the tracking art must be correspondingly advanced at all levels. If this is not done, a stage will be reached where missile contractors will find it necessary to abandon the use of external data for the evaluation of the finer points of missile performance and, instead, to place their reliance solely upon accurately determined impact points.

The primary reason for these conflicting trajectories is that practically all data acquisition systems possess undetermined bias or systematic errors. To illustrate this point consider typical azimuth, elevation, and range (A^m, E^m, R^m) data versus time, T^m, from a radar on a missile launch. The superscript m represents measured quantities. Let us assume that we actually had some means of knowing true or absolute data (A^t, E^t, R^t), that is, let us assume we knew exactly where the missile was at any time, T^t. The superscript t represents true or absolute quantities. A comparison of these data will reveal that the data acquired by the radar and the true data differs from one another. This difference is a consequence of the random and systematic errors of the radar. If there were some means on an actual missile test of determining various control points such as the assumed true data on the above example, then it should be possible to eliminate the systematic errors from the data. Moreover, if the systematic errors are long termed, the accumulation of data from several tests should permit calibration of the radar.

The procedures utilized at the Atlantic Missile Range for calibrating the systematic errors in highly precise electronic tracking systems are presented in this paper. Among the systems are the Mark II AZUSA which is designed to measure absolute accuracies of 1 part in 100,000 and the MISTRAM system which has design specifications for measuring absolute accuracies of 1 part in 1,000,000.

In view of the foregoing, it is clear that the problem of resolving conflicting trajectories, which in turn is ultimately one of instrument calibration, must be solved on any range which is charged with the responsibility of producing accurate data for the precise evaluation of missile performance. The effective calibration of tracking systems is not a task for the inexperienced, since capability in this highly specialized area cannot be acquired in a short period of time. It requires a degree of sophistication which goes far beyond that necessary for the computed reduction of data. It demands an intimate knowledge of the data acquisition instrument, including the environment and manner in which the instrument is used; the data recording media associated with the instrumentation; the data translating equipment which translates the electrical or optical image to digital computer language; the mathematical techniques for reducing the data to intelligible information, and the handling procedures employed by personnel in processing the data. It is important to note that data reduction techniques reach full maturity only when all of the data acquisition instrumentation from which data are reduced have been analyzed for all sources of error and these errors have been calibrated.

NATURE OF INSTRUMENTAL ERRORS

Instrumental errors may be classified as random and systematic. Random errors are chance variables and are individually unpredictable. They may be characterized only in statistical terms and may be effectively reduced only by statistical means. Systematic (or bias) errors, on the other hand, are generally continuous functions of various physical parameters and are subject to calibration and to correction.

In the discussion of the relations between random and systematic error it is convenient to introduce the concept of the "Bias Factor" of a system. This is defined as the ratio of the bias or systematic error in a typical observation to the standard deviation

of the observation (the standard deviation, of course, is a measure of the purely random error). For example, cinetheodolites have typical standard deviations of the measured angles which range from 5 to 10 seconds of arc, while typical systematic errors range from 20 to 40 seconds of arc. Hence, the bias factor for cinetheodolite data may range from 2 to 8. The Mark I AZUSA system has typical standard deviations of 3 to 5 parts per million in angle and 5 feet in range while typical systematic errors range from 25 to 200 parts per million in angle and 10 to 50 feet in range. Hence, the bias factor for the Mark I AZUSA may range from 5 to 75 for angle and from 2 to 10 for range. Practically all tracking systems have bias factors greatly exceeding unity. It is this which leads to conflicting trajectories.

Accuracywise, a system may be regarded as completely satisfactory, within its inherent capability, only when the systematic errors in the basic observations have been suppressed to a level appreciably below that of the typical random error--that is to say, only when the bias factor of the system has been reduced substantially less than unity. Only then is it possible to provide a positive and well defined answer to the ever recurring question of "How good are the final data."

CALIBRATION OF TRACKING SYSTEMS

The adequate determination of the random errors affecting trajectory data usually presents no difficulties and may be ordinarily accomplished by established procedures of time series analysis. On the other hand, the problem of evaluating the system errors in a given tracking system is considerably more difficult unless a suitable standard for comparison is available. Time series analysis reveals nothing about systematic errors.

The actual calibration of a tracking instrument involves a combination of approaches. The first requirement is that a suitable error model be formulated mathematically. Reasonable errors in each of the model parameters are introduced and the resultant effects on the data are calculated. With a radar, for instance, the systematic error in the angles may be expressed in terms of such parameters as misalewl, index error, collimation error, etc. The process of calibration consists essentially of the determination of the unknown parameters appearing in the error model. This is accomplished in practice by fitting the error model to a set of known systematic errors corresponding to various instrumental settings. The necessary systematic errors are obtained for the most part by means of laboratory evaluation of the components which can reasonably be expected to remain stable in the field, and by the comparison of observations resulting from different instrumentation systems.

At the Atlantic Missile Range the ballistic camera is used as the standard for comparison of other range instrumentation. The ballistic camera is recognized as the only system of high precision which can consistently produce data having a bias factor of less than one. Consequently, it is the only system for which meaningful and precise statements can be made concerning the absolute accuracy of the final product. It is therefore, an excellent standard for calibrating the systematic errors in electronic systems.

CALIBRATION OF THE MARK II AZUSA

As an illustration of the calibration procedure, I would like to outline the mathematical approach and test design for the evaluation and calibration of the Mark II AZUSA. The general antenna configuration of the system is given in Figure 1.

The site transmits a carrier frequency of approximately 5060.2 mc frequency modulated with three ranging frequencies. The airborne transponder locks on to this signal, offsets the carrier by approximately 60.2 mc and transmits this signal to the site. The site transmitter frequency is adjusted so that the received frequency at the site is always 5000 mc.

AZUSA range data is obtained by phase comparing the transmitted ranging frequency with the received ranging frequency. Where the transponder is less than one-half wavelength of the measuring frequency from the site, the measured phase angle is proportional to range.

When the transponder is moving relative to the site, the Doppler Effect will appear as changes in phase measurement. These appear as counts in the data units so the range measurement may be considered as an initially determined value plus a summation of range increments.

Coherent carrier range is based on the same principle but is derived from the 5000 mc carrier frequency.

Cosine data is obtained by phase comparing the 5000 mc signals received by two baseline antennas. The measurement obtained is a range difference, which can be related to a direction cosine from the site.

An error analysis of the Mark II AZUSA has resulted in the following general mathematical model of the system.

$$l_i^t = l_i^m + v_{l_i} + \beta_l$$
$$= l_i^m + v_{l_i} + a_0 + a_1 l_i^m + a_2 m_i^m + a_3 n_i^m$$

$$m_i^t = m_i^m + v_{m_i} + \beta_m$$
$$= m_i^m + v_{m_i} + b_0 + b_1 l_i^m + b_2 m_i^m + b_3 n_i^m$$

$$r_i^t = r_i^m + v_{r_i} + \beta_r$$
$$= r_i^m + v_{r_i} + c_0 + c_1 r_i^m$$

$$\dot{l}_i^t = \dot{l}_i^m + v_{\dot{l}_i} + \beta_{\dot{l}}$$
$$= \dot{l}_i^m + v_{\dot{r}_i} + d_0 \dot{l}_i^m + d_1 \dot{m}_i^m + d_2 \dot{n}_i^m$$

$$\dot{m}_0^t = \dot{m}_i^m + v_{\dot{m}_i} + \beta_{\dot{m}}$$
$$= \dot{m}_i^m + v_{\dot{m}_i} + e_0 f_i^m + e_1 \dot{m}_i^m + e_2 \dot{l}_i^m$$

$$\dot{r}_i^t = \dot{r}_i^m + v_{\dot{r}_i} + \beta_{\dot{r}}$$
$$= \dot{r}_i^m + v_{\dot{r}_i} + e_1 \dot{r}_i^m$$

Where the β quantities represent the bias errors of each of the six AZUSA parameters.

The V quantities represent the random errors of each of the six AZUSA parameters,

The subscripted quantities a_0, a_1, etc. are unknown coefficients of the measured quantities,

and $n_1^2 = 1 - l_1^2 - m_1^2$.

Each of the a_0, a_1, etc. terms in the expression represent bias errors of a particular kind. The a_0 term in β_1 represents the zero-set errors of the l cosine measurement. The remaining three terms in β_1 , i.e., a_{1l}, a_{2m} and a_{3n} represent l cosine errors caused by baseline errors of length, azimuth and tilt.

Special tests are conducted at night in which the instrumentation being calibrated and the ballistic cameras simultaneously track a test aircraft over a predetermined flight path. A typical flight path is shown in Figure 2. The aircraft carries a flashing light which produces pin-point exposures on the ballistic camera plate. The time of these exposures is recorded through the medium of Range Timing. Thus accurate time-position information is obtained from the ballistic camera. The instrumentation under test tracks the aircraft in its normal mode yielding another set of time position data.

The reduction of AZUSA and Ballistic Camera data proceed independently and follow standard data reduction procedures. Special attention is given to the problem of performing refraction corrections, the ray tracing method (i.e., numerical integration based upon observed index of refraction profiles) being employed in both cases.

The final output of the AZUSA reduction consists of t, l, m, r, $\dot{l}$, $\dot{m}$, $\dot{r}$ corrected for all known effects.

The ballistic camera coordinates are corrected for the displacement between the strobe lamp and the AZUSA antenna. Then, the data are converted to direction cosines and range (l^t, m^t, r^t) referenced to AZUSA. In the comparison these are regarded as being the 'true' values. The covariance matrices of the computed l^t, m^t, r^t values are also computed.

The discrepancies between the Ballistic Camera data and the corresponding AZUSA data are computed:

$$\Delta l = l^m - l^t$$

$$\Delta m = m^m - m^t$$

$$\Delta r = r^m - r^t$$

Since the uncertainty in the refraction corrections may be significant for data runs at low elevation angles, the azimuth discrepancies, which are unaffected by normal refraction, are also computed:

$$\Delta A = A^m - A^t$$

where

$$\tan A^m = \frac{l^m}{m^m}, \quad \tan A^t = \frac{l^t}{m^t}$$

An error model of the following form is then fitted to the observed discrepancies:

1. $\Delta l = a_0 + a_1 l + a_2 m + a_3 n$

2. $\Delta m = b_0 + b_1 l + b_2 m - b_3 n$

3. $\Delta r = c_0 + c_1 r$

4. $\Delta A = \frac{a_1 m}{\beta} + \frac{a_1 l m}{\beta} + \frac{a_2 m^2}{\beta} + \frac{a_3 m r}{\beta} - \frac{b_1 l}{\beta} - \frac{b_1 l^2}{\beta} - \frac{b_2 l m}{\beta} - \frac{b_3 l r}{\beta}$

The $\dot{l}$, $\dot{m}$, $\dot{r}$ data produced by AZUSA are integrated numerically and are likewise compared with ballistic camera data.

For a given data run either equations 1, 2 and 3 are employed in the adjustment or else equations 3 and 4, depending upon whether the run is of high or low elevation.

Since the errors in the discrepancies Δl, Δm, Δr, ΔA are correlated, all four are carried through a common adjustment. This leads to the formation of a 10 x 10 system of normal equations, the solution of which yields the calibration parameters a_0, a_1, a_2, a_3, b_0, b_1, b_2, b_3, c_0, c_1. These are then employed in equations 1, 2, 3 and 4 to obtain the computed discrepancies $\Delta\hat{l}$, $\Delta\hat{m}$, $\Delta\hat{r}$, $\hat{\Delta A}$. The residual discrepancies are then computed from:

$$\delta l = \Delta l - \Delta\hat{l}$$

$$\delta m = \Delta m - \Delta\hat{m}$$

$$\delta r = \Delta r - \Delta\hat{r}$$

$$\delta A = \Delta A - \hat{\Delta A}$$

It is the residual discrepancies which determine whether or not the system is capable of meeting specifications.

The ten calibration parameters a_0, ..., c_1 are compared from test to test, both individually and collectively, to determine whether any significant variation exists. The covariance matrix of each group of calibration parameters is obtained from the inverse of the normal equations.

The stability of the calibration parameters from test to test will reflect the physical stability of the Mark II AZUSA System.

CALIBRATION BY "BEST ESTIMATE OF TRAJECTORY" APPROACH

In addition to the above described calibration techniques where an accurate comparison standard exists, another approach which exploits the redundancy of data from several instrumentation systems is being utilized at the Atlantic Missile Range. The approach has been called "The Best Estimate of Trajectory Utilising Instrument Error Models." This technique utilises the error model equations and the fact that systematic errors change slowly with time.

To illustrate the principle involved in the approach, consider a hypothetical problem involving data simultaneously acquired by AZUSA, Radar, and two cinetheodolites. The observations may be denoted by:

AZUSA:	l, m, r
Radar	A, E, R
Theodolite No. 1:	A^1, E^1
Theodolite No. 2:	A^2, E^2

Thus, each trajectory point will give rise to 10 observations.

If it is assumed for the moment, that each type of observation is affected by constant bias, then observations for the i^{th} point will be interrelated by the following equations:

$$l_i + v_{l_i} + \beta_l = l_i^t = f_1(x_i, y_i, z_i)$$
$$m_i + v_{m_i} + \beta_m = m_i^t = f_2(x_i, y_i, z_i) \quad \text{Mark I AZUSA}$$
$$r_i + v_{r_i} + \beta_r = r_i^t = f_3(x_i, y_i, z_i)$$

$$A_i + v_{a_i} + \beta_A = A_i^t = f_4(x_i, y_i, z_i)$$
$$E_i + v_{e_i} + \beta_E = E_i^t = f_5(x_i, y_i, z_i) \quad \text{FPS-16}$$
$$R_i + v_{r_i} + \beta_R = R_i^t = f_6(x_i, y_i, z_i)$$

$$A_i^{(1)} + v_{a_i}^{(1)} + \beta_A^{(1)} = A_i^{(1)t} = f_7(x_i, y_i, z_i) \quad \text{THEODOLITE NO. 1}$$
$$E_i^{(1)} + v_{e_i}^{(1)} + \beta_E^{(1)} = E_i^{(1)t} = f_8(x_i, y_i, z_i)$$

$$A_i^{(2)} + v_{a_i}^{(2)} + \beta_A^{(2)} = A_i^{(2)t} = f_9(x_i, y_i, z_i) \quad \text{THEODOLITE NO. 2}$$
$$E_i^{(2)} + v_{e_i}^{(2)} + \beta_E^{(2)} = E_i^{(2)t} = f_{10}(x_i, y_i, z_i)$$

in which the v's are random errors, x, y, z, are the coordinates of the observed trajectory point and the β's are the observational biases. It will be noted that the β's are the same for all trajectory points considered and hence are independent of i. It is this persistence of bias which makes a solution feasible. It is seen that a total of n trajectory points will lead to a system of $10n$ equations in $3n + 10$ unknowns (a new x, y, z will be unknown for each trajectory point). Therefore, in the absence of random measuring errors, just two trajectory points would provide more than enough equations to solve for the 10 biases and x, y, z's. In practice, of course, the logical procedure would involve a least squares solution utilizing all available data points simultaneously.

In a numerical study designed to test the feasibility of the solution it was assumed that data were acquired from the above systems on a theoretical IRBM trajectory. For computational ease, seventeen equally timed points over 170 seconds of data were selected. The true observations, i.e., l_i^t, m_i^t, etc., were determined for each instrument from the theoretical trajectory. Random errors, v_{l_i}, v_{m_i}, etc., from a table of random numbers were added. Typical or "true" bias errors were then added to the data as shown in Table 1 below. The computed biases from the BRT computation and their 2σ estimates were then obtained and are also tabulated.

True Biases	Computed Biases $\pm 2\sigma$	
$\beta_l = -.000098$	$-.000129 \pm .000026$	MARK I AZUSA
$\beta_m = .000059$	$+.000069 \pm .000016$	
$\beta_r = 2.00$	2.32 ± 4.10	
$\beta_A = .0000306$	$.000291 \pm .000028$	FPS-16 RADAR
$\beta_E = -.000136$	$-.000105 \pm .000036$	
$\beta_R = 132.0$	130.6 ± 5.5	
$\beta_A^{(1)} = -.000167$	$-.000131 \pm .000030$	THEODOLITE NO. 1
$\beta_E^{(1)} = .000032$	$-.000008 \pm .000040$	
$\beta_A^{(2)} = .000127$	$+.000126 \pm .000025$	THEODOLITE NO. 2
$\beta_E^{(2)} = -.000037$	$-.000004 \pm .000036$	

TABLE 1

It is interesting to note that only β_l and $\beta_A^{(1)}$ are outside of the 2σ limits and that they are just outside of these limits. Moreover the accuracy of determining the Betas would increase by the square root of the number of data points utilized. Thus, if 170 points had been used, the 2σ limits on β_l would have been $\pm .000026/3.2$ or $\pm .000008$.

In the numerical example, the Betas were assumed to be constant. However, the Betas could be analytical expressions of the error models for each instrument. The solution for the parameters, a_0, a_1, etc. on each test for which Mark II AZUSA is utilized will yield continuous and long term information on the stability, calibration, and overall performance of the system.

Calibrating an electronic system by the "Best Estimate of Trajectory" approach would, at first sight, appear to be prohibitively difficult from a computational standpoint. A problem involving p error parameters and n trajectory points would lead to a system of normal equations of order $(3n + p) \times (3n + p)$. It turns out, however, that the normal equation coefficient matrix is so patterned with zeroes that its inversion is readily accomplished by partitioning; the largest matrix which actually need be inverted is of order $p \times p$. Moreover, the computations may be so arranged that the data for each trajectory point may be processed independently. The amount of internal computer storage and the needed computer capacity are therefore essentially independent of the number of data points but depend rather upon the number of error model parameters, p. The computing time increases only linearly with n. More generally, the computing time increases as np^2, (where $n \gg p$), so that by doubling the number of error parameters one quadruples the overall computing time.

Finally, the calibration of the MISTRAM system which has design specifications for

measuring 1 part in 1,000,000 will follow procedures similar to those presented above. Of course, the calibration of MISTRAM will require considerable more care. For example, the accuracy requirements from the comparison standard, i.e. ballistic camera data, will naturally be more stringent. Consequently, AMR is procuring 600 mm focal length ballistic cameras for obtaining more accurate data for comparison. Moreover, critical locations are being resurveyed to obtain an order of magnitude improvement in accuracy and improved methods of measuring tropospheric refractive index are being explored.

SUMMARY

Methods employed at the Atlantic Missile Range for calibration of electronic tracking systems were discussed.

In particular, discussions of two approaches to instrument calibration were presented. The first of these approaches outlined the utilization of ballistic camera data as a standard for comparison and consequently, as a means for calibration. Specifically, attention was focused on the utilization of this approach for calibrating the Mark II AZUSA, which is designed to measure 1 part in 100,000.

The second of these approaches outlined the "Best Estimate of Trajectory Utilizing Instrument Error Models" as a means for calibration of the instrumentation system. A numerical example illustrating this approach was also presented.

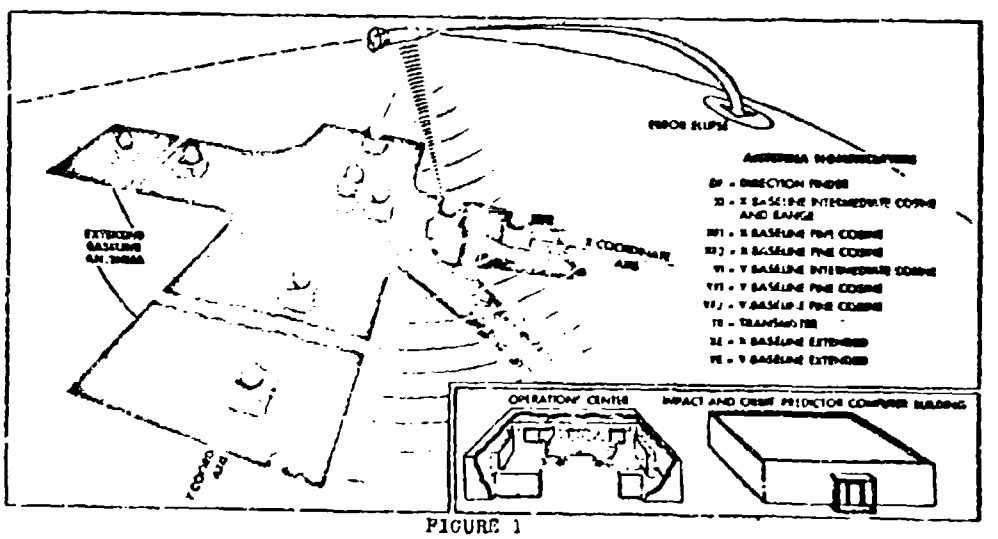
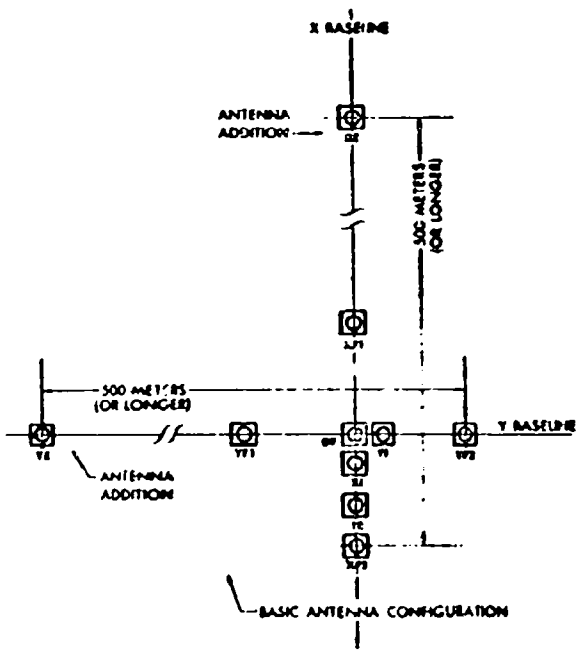


FIGURE 1



Extended Baseline Azusa System

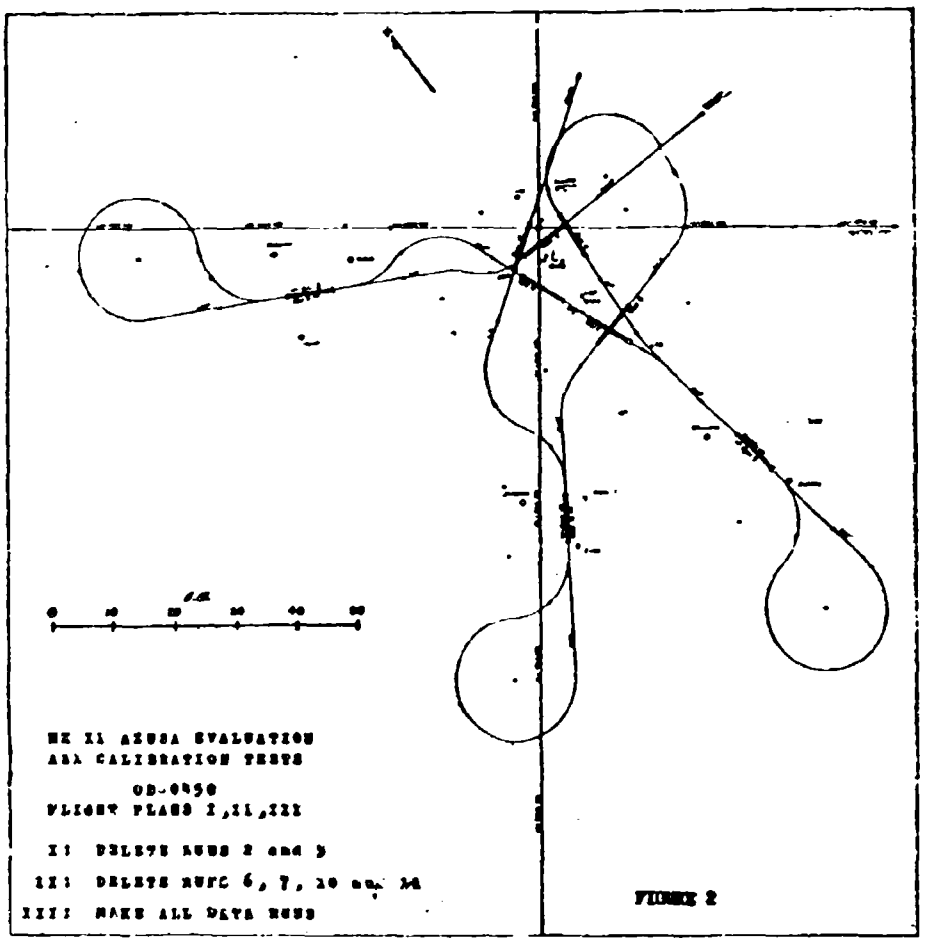

NK XI AZUSA EVALUATION
ARA CALIBRATION TESTS
OB-0850
FLIGHT PLANS I,II,III
XI: DELETE RUNS 2 and 5
XII: DELETE RUNS 6, 7, 10 and 14
XIII: MAKE ALL DATA RUNS

FIGURE 2

COMMUNICATIONS BY RE-RADIATION FROM CHAFF

By: Mr. R. V. Hess, Signal Communications Department, USAEPG, Fort Huachuca, Arizona

Modern weapons will have considerable impact on communication systems of the future. Increased dispersion of our forces will require more communication for proper command control and will necessitate communication over greater distances. Mobility of future systems will have to be improved and the time required for a system to react to a situation will have to be reduced. Of these demands, the requirement for increased distance is the most difficult to obtain since it is here that nature seems to conspire against us.

At the present time we use three basic methods to obtain reliable trans-horizon communication at the higher frequencies. One method is to place relays at each successive radio horizon until we have covered the distance required. This is the well known radio-relay technique.

The second method is to reflect signals off the ionized trails of meteors.

The third method is the new tropo-scatter technique wherein the large path losses involved are countered by brute force methods.

A fourth possible method of maintaining communications over distances well beyond horizon is by re-radiation or forward scatter from radar type chaff. The scattering pattern of an ensemble of chaff dipoles was analyzed by Dr. R. A. Hessemer, Jr. of the Applied Research Laboratory of the University of Arizona under contract DA-36-039-SC-80146.[1]

In his work Dr. Hessemer found that the radiation pattern of the chaff ensemble is an oblate spheroid, with the minor axis parallel to the antenna and one third that of the major axis. For spherically dispersed, random oriented chaff the maximum cross sectional area is $0.16\ N\lambda^2$ where N is the number of half wave dipoles and λ is wave length. If the spherically dispersed chaff is oriented randomly in a horizontal plane the chaff cross section area becomes $0.29\ N\lambda^2$. Fluttering and dispersion of the chaff causes the amplitude of the received signal to have a Rayleigh time distribution. If the chaff is constricted so that it falls essentially vertically polarized but is oriented randomly within a solid cone, the effective cross-sectional area is also a function of the cone angle and the angle between the normal from the chaff ensemble and the electric field from the antenna. This relation is shown in Figure 1.

The analytical work of Dr. Hessemer assumes that the chaff ensemble is composed of half-wave dipoles. This is only true if the chaff particles remain essentially rigid. If the dipoles bend they will act as though they had an electrical length short compared to a half-wave. In this case their re-radiation efficiency will be seriously degraded. At frequencies where the chaff geometry becomes such that rigidity cannot be maintained, it should be possible to shape the chaff so that it will fall essentially vertically polarized within some solid angle. Extremely short dipoles, 2 or 3 inches in length, tend to fall predominantly in a horizontally polarized position.[2] Such chaff falls at a rate of about 2/3 that of vertical dipoles. Chaff 10 to 12 inches in length seems to assume a truly random orientation. The rate at which the chaff will fall is a function of a number of variables and is not too well known. Rates as low as 100 to 300 feet per minute should be achievable.

The path loss of a system involving chaff is the sum of the free space loss between the transmitter and the chaff ensemble, the chaff absorption and re-radiation loss, the free space loss between the chaff and the receiver and the effective cross section of the receiver isotropic antenna. This can be expressed as

$$L(db) = 10\log + \pi D_T^2 - 10\log \bar{\sigma} + 10\log 4\pi D_R^2 - 10\log \frac{\lambda^2}{4\pi}$$

$$= 10\log (4\pi D_T D_R)^2 - 10\log \bar{\sigma} - 10\log \frac{\lambda^2}{4\pi}$$

D_T is the distance from the near antenna to the chaff cloud. D_R is the distance from the chaff cloud to the far antenna. λ is wave length. $\bar{\sigma}$ is the average chaff cross section area $= f(N, \lambda^2)$ where N is the number of chaff dipoles. For a constant weight of chaff and a constant length to width to thickness ratio, N varies inversely with λ^3. $\therefore \bar{\sigma} = f(\frac{1}{\lambda})$
The total path loss therefore varies directly with λ.

The maximum allowable loss is a function of the antenna gains as well as path loss. Antenna gains, for a given size and weight, vary inversely with wave length. Within limits, the maximum allowable system loss for a system involving re-radiation from chaff therefore varies directly with frequency. In investigating system performance a frequency of 2 kmc was assumed. Due to the short length of the half wave dipoles at this frequency chaff will fall horizontally polarized and therefore be random in only the x-y plane rather than in all three planes. This will provide a 3db decrease in path loss over purely random chaff. The 2 kmc band is also assigned to the military and is lightly populated.

It will be noted that the path loss varies directly as the product of the two distances involved. For a given total distance (D_T / D_R) the product will be a minimum when either D_T or D_R is

[1] Forward Scattering with Radar Chaff, Dr. R. A. Hessemer Jr. submitted for publication in Transactions of the Professional Group On Antennas and Propagation.

[2] Kuiper G. P., "A Study of Chaff Echoes at 515 MC", RDU. Harvard University Report 411-73, Dec 19, 1943.

a minimum. Thus the path loss will be a minimum if the chaff is emplaced directly over either the transmitter or the receiver.

Figure 2 is a plot showing the relative path loss as a function of the slant distance from the chaff ensemble to the near antenna. It will be noted that the system loss is very sensitive to changes in this distance. A change from 5,000 feet to 10,000 feet at the near terminal has the same effect as a change in distance of 50 to 100 miles from the chaff ensemble to the far terminal. In calculating system losses it must be kept in mind that the chaff is always falling, and also is moving horizontally unless there is a dead calm. The slant distance from the chaff ensemble to the near antenna is therefore a function of time.

Figure 3 is a plot of path loss as a function of communication range for a 2 kmc system. In this Figure the chaff height is a constant, equal to that needed to provide a line-of-sight path to the other terminal plus a margin of 3000 feet to allow for chaff vertical displacement with time. The cross hatched area includes an allowance for increased path loss due to horizontal displacement with time. The path loss of a system employing tropospheric scatter propagation is included as a reference.

It is evident from this figure that the use of chaff as a scatter medium offers a definite advantage over the tropospheric scatter mode for long distance communications. In investigating such an application the following system constraints are assumed:

a. A communication range of 250 miles is desired.

b. The system shall be capable of supporting four voice channels plus an order wire.

c. System reliability shall be 99.6% or better.

d. System must be highly mobile and therefore capable of being transported in a 2 1/2 ton 4x4 truck.

e. To reduce vulnerability and reaction time antenna diameters must not be greater than 8 feet at the forward station and 15 feet at the rear station.

For 99.6% reliability a fade margin of 15db must be allowed. The noise power in the receiver IF for an IF bandwidth of 100 kcs and a 4db receiver noise figure = -204 / 50 / 4 = -150dbw. The minimum received power which will permit a 15db fade margin over a 10db threshold = -150 / 15 / 10 = -125dbw. The path loss of a 250 mile chaff re-radiation system is 214db. The effective radiated power must therefore be 214 - 125 = 89dbw. Assuming an 8 foot dish at the forward terminal and a 15 foot dish at the rear terminal, a total gain of 31.5 / 37 = 68.5db is obtained. With this antenna gain a transmitter power of 20.5db is needed to achieve the 89dbw effective radiated power required. 20.5dbw is 112 watts, a transmitter which is actually only "Jeep" sized. If smaller antenna dishes were desired and a 1 kw transmitter in a 2 1/2 ton 4x4 vehicle was acceptable, the sum of the antenna gains could be reduced to 89 - 30 = 59db. This total antenna gain can be obtained using a 6 foot dish at the forward terminal and an 8 foot dish at the rear terminal. Each chaff emplacement would provide over 20 minutes of communication half the time and over 10 minutes communication 99% of the time.

A tropospheric scatter circuit, to obtain the same communication, would require an effective radiated power 20db greater since the basic path loss using this propagation mode is 20db more. Even with 15 foot dishes at both ends a transmitter power output of over 3 kw would be required.

As a single channel voice set a 100 watt transmitter, employing the chaff re-radiation mode, could provide communications up to 300 miles with a reliability of 99%. In this system a 5 foot and an 8 foot dish are hypothesized at the forward and rear terminals respectively.

Referring again to Figure 3 it would appear that tropo scatter would be the preferred propagation mode for circuits of less than 80 miles. These curves though, are based on a horizon angle (Ψ) sum of 0 degrees. For horizon angle sums greater than 0 degrees, the advantage of the chaff mode is greatly enhanced, particularly on relatively short paths. Figure 4 shows the loss due to horizontal angles for several distances. Losses due to elevated horizons are independent of distance in the chaff mode. Even if the horizon angle sum is only 1 degree the chaff mode advantage is 19 - 3.4 = 15.6db for a 50 mile circuit. Thus basic path losses of 179db and 189.5db for 50 mile tropo and chaff modes respectively become 198db and 193db when Ψ is 1 degree. In addition, for a given reliability, the fade margin required becomes quite large for shorter communication ranges employing tropospheric scatter. Figure 5 shows reliability as a function of fade margin for various path lengths. For 99% reliability the fade margin for a 50 mile circuit employing dual diversity is 26.5db as compared to 13.5db for a 200 mile circuit. Most of this increase in fading margin with decrease in distance seems to be due to the scatter volume being at a lower altitude where it is more subject to atmospheric changes. When employing forward scatter from chaff the only factor in long term fading is that due to changes in atmospheric conditions between the chaff ensemble and the distant antenna. The fading margin required for a given reliability should be considerably less for the chaff mode than for tropospheric scatter mode for the shorter communication distances. It is therefore reasonable to expect about the same performance from the two systems for horizon angle sums of 0 degrees and greatly improve performance from the chaff system if elevated horizon angles are involved.

Let's investigate a single channel voice system having a reliability of 99% in which the forward terminal must be Jeep mounted with an antenna not over 3 feet in diameter. It is assumed that a rear terminal antenna size of 6 feet is permissible and communications over paths where the sum of the horizon angles are 1 degree or less is desired. Assuming a frequency modulated system with a receiver noise figure of 4db and an IF bandwidth of 25kc, the noise in the receiver IF

is -204 / 44 / 4 = -156db. The minimum received power for a threshold of 10db and a fade margin of 12db is - 156 / 10 / 12 = -134db. A 40 watt transmitter is capable of Jeep mounting. The effective radiated power of such a transmitter employing a 3 foot dish at one end and a 6 foot dish at the other end would be 16 / 23 / 29 or 68dbw. The maximum path loss would therefore be 68 / 134 - 4 = 193db. Such a system should be capable of providing communication over distances in excess of 100 miles. One application for such a system might be in long range reconnaissance patrol communications.

As mentioned earlier both the chaff height and its horizontal position vary with time. Its rate of fall is primarily a function of chaff geometry while its horizontal displacement is primarily a function of wind velocity. World wide the wind velocity is less than 10 miles per hour 50% of the time and less than 30 miles per hour 99% of the time. It should be possible to design 2kmc dipoles so that they would fall horizontally polarized at a rate of about 150 feet per minute. In this paper a fall rate of 166 2/3 feet per minute was assumed since it simplified calculations. Under average conditions the chaff would therefore fall 1000 feet and be displaced horizontally one mile in 6 minutes. Figure 6 shows the relative path loss as a function of time for various initial chaff heights under these average conditions. If the initial chaff emplacement was 3000 feet above that necessary to provide line of sight to the distant antenna a single chaff emplacement would provide communications for more than 18 minutes half the time.

Under higher wind velocity conditions, lateral rather than vertical displacement would be the liriting factor. Even under these conditions, it should be possible to maintain communication for more than eight minutes for each chaff emplacement 99% of the time.

If this technique could be used to extend the range of the VHF tactical FM equipment, further applications would be possible. At these low frequencies the chaff cannot approach horizontal polarization without bending. It would therefore be necessary to so design the chaff that it would tend to fall vertically polarized. Experiments are necessary to determine just what can be accomplished along these lines. For purposes of our analysis, it is assumed that it is feasible to so design chaff that it will fall spherically dispersed and randomly distributed within a conical angle of 20 degrees about a vertical. Under these conditions an additional factor is introduced, that of the coupling between the antenna and the chaff ensemble. As shown in Figure 1, path loss with respect to that of purely random oriented chaff will vary not only with the conical angle but also with the angle between the electric field of the antenna and a normal to the ground. For the 20° conical angle assumed, an additional path loss of 8.7db is experienced if the chaff is directly overhead. (θ_3 - 90°) while an ensemble gain of 6.5db is obtained if the angle is 0 degrees. If the chaff is emplaced initially so that this transmitter angle is always less than 70° during the useful life of the chaff cloud, this coupling loss can be minimized.

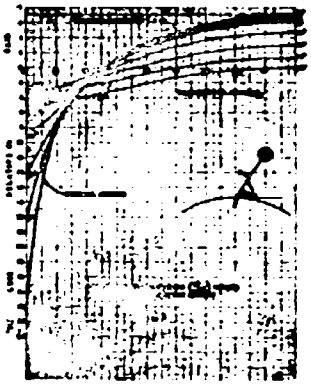
Figure 7 gives the path loss for communications ranges of 20 to 50 miles with a useful life expectancy of 5 to 10 minutes. In this figure 200 dipoles randomly distributed in a 20° cone were assumed. Chaff height was taken as that necessary to provide a line of sight path to the far antenna plus 2500 feet for a vertical displacement with time. No allowance was made for shrouded horizons. The cross hatched area takes into consideration gains and losses due to horizontal displacement and antenna-chaff coupling. The useful life of each chaff ensemble can be extended if the increased loss due to horizontal displacement can be tolerated.

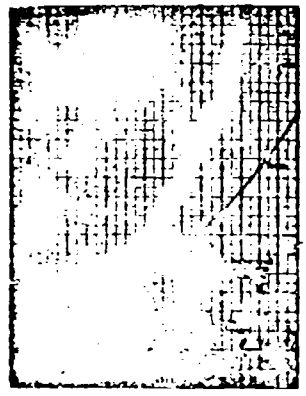
FM equipments such as the AN/VRC-12 have a noise power in the IF of -204 / 454 / 6 = -152.6 dbw. Allowing for a 10db threshold margin results in a minimum received power of -142.6dbw. The effective radiated power is 18.8dbw thus giving a maximum path loss of 142.6 / 18.8 = 161db. This equipment could thus be used to communicate over a 50 mile diameter circle with a fade margin of 23db, and over a 100 mile diameter circle with a fade margin of 13db.

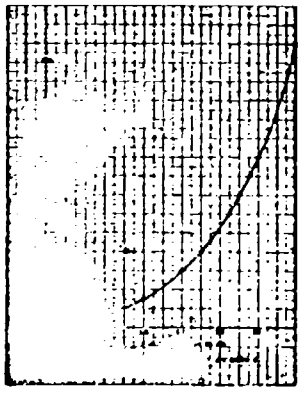
One application for such a system might be a general broadcast facility such as the Navy "Fox" system. This facility could be used to disseminate general information or to broadcast warnings over large areas. The auxiliary receiver of radio set AN/PRC-25 provides a ready means to extend such communication down to the platoon level. Since one control station can cover an area of thousands of square miles, economically, the central station transmitter need not be limited to the 25-35 watts of the VRC-12. A one Kilowatt central station could be used to serve an entire army area if desired.

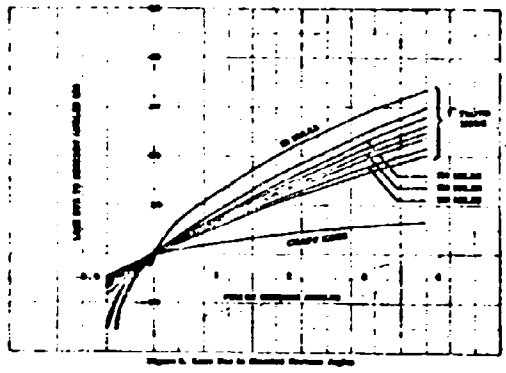
A second application might be as an emergency capability when communications using natural means has been lost due to distance or terrain. The feasibility of this application would hinge on the development of an inexpensive means of emplacing chaff at altitudes of 4 to 5 thousand feet. A small rocket type device might meet this requirement.

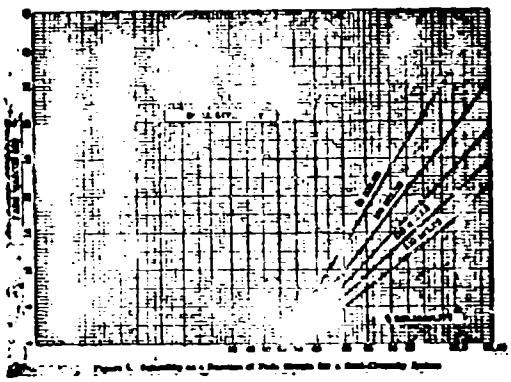
Work to date has been entirely theoretical in both the analysis of the basic technique by Dr. Kesmarky and in the system feasibility study presented herein. An experimental program is now needed to further investigate applications of the technique and problems in its implementation. Further information on the rate of descent and on dispersion characteristics of chaff dipoles is required. Methods for the emplacement of the chaff ensemble in its optimum point in space with the chaff dipoles so oriented as to reduce system losses to a minimum need to be developed. Analytical studies to date have indicated the desirability of these further investigations.

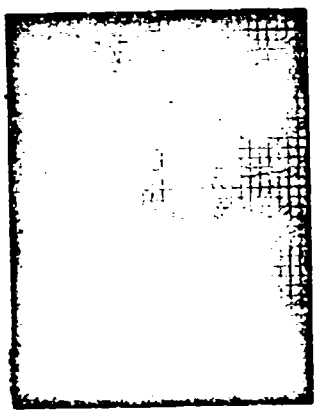


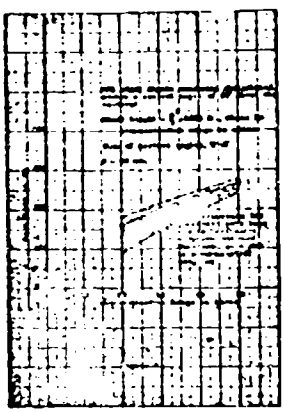












FLIGHT MEASUREMENTS ON THE JUPITER R&D MISSILE

By: C. T. E. Paludan
Army Ballistic Missile Agency

Abstract

An introduction to the organization at ABMA with emphasis on the Measuring Systems Section, and to the JUPITER Missile is given. A statement of the philosophy behind a measuring system is made. A brief description of the module system is given. Individual measurements are discussed in some detail, with emphasis on methods and equipment. Mention is made of special measurements, including those on monkeys and on NASA projects. A brief summary of results concludes the discussion.

Discussion

Measurements made during flights of missiles developed by the Army Ballistic Missile Agency are the responsibility of the Measuring System Section. The position of this Section in the organization can be seen on the chart in Figure 1. The Section consists of several Units, each responsible for a particular area of measurement.

In order to present the measuring scheme for a large ballistic missile, the R&D version of the JUPITER IRBM has been chosen. It is shown in Figure 2 at the moment of liftoff from its launching pad at the Atlantic Missile Range.

The background philosophy behind measurements on a missile stems from a desire to obtain design data for future missile flights. This means that each missile flight test could possibly dictate minor design changes and indicate problem areas for concentrated study on the next test flight.

The implications of this are broad, but one which has a profound effect is that the measuring system must have the utmost flexibility. Furthermore, a large number of measurements are required by this philosophy. These two factors conflict with one another.

The formulation of measuring programs has presented certain problems. Requests for measurements on the missile generally come from Structures and Mechanics Laboratory, Aeroballistics Laboratory, and Guidance and Control Laboratory. These are assigned various orders of priority by the originating laboratory's coordinator. After receipt, contacts are made with measuring specialists to determine feasibility and availability of hardware. There is then a procedure of obtaining electrical power, electrical networks, telemetering channels, mounting and installation coordination, etc.

The basic aims of ABMA's measuring programs have been concerned with the development of the missiles themselves. Purely "scientific" research measurements have had secondary importance to that. We believe we have been successful in providing this country with information vital to the defense program. At the same time we have provided a framework for measurements on projects with purely peaceful purposes.

Figure 3 is a condensed version of a typical measuring program for JUPITER. These will be discussed in detail later.

Two factors affect the selection of components for flight measurements: (1) they must either produce or be adaptable to produce the correct electrical outputs for telemetering, and (2) they must function properly in the severe environment encountered. The first of these conditions in many cases necessitates electronic signal conditioning. The second implies a rigorous testing program of many alternate components.

Before going into a discussion of individual measurements, I would like to present some general discussion of signal conditioning.

In the early REDSTONE Missiles the necessary signal conditioning was accomplished by a non-homogeneous collection of electronic circuits housed in containers of various sizes. A more homogeneous group of devices developed from these later in the REDSTONE program. They were put to more extensive use on early JUPITERS (Fig. 4).

A more sophisticated system evolved directly from the earlier units, the principal change being the packaging. As can be seen in Figure 5, a modular concept was used. This permitted rapid exchange of signal conditioning circuits in case of additions, deletions, modifications, or malfunctions. Several basic modules were and are available: a narrow band D.C. amplifier, a wide-band D.C. amplifier, a carrier amplifier, an A.C. amplifier, and a power supply. In addition to these, there are a few special purpose modules such as a flowmeter converter, a servosweeper, and an inverter frequency adapter. Figure 6 outlines the basic functions of the module.

Four types of pressure transducers were commonly used. The most popular was the potentiometer type. With it, the desired electrical output for telemetering (0-to 5 v.d.c.) was obtained without amplification. It had limited frequency response. For higher frequency response, smaller size and smaller line volume, the strain-gage type was sometimes used. Its output required amplification. For greater accuracy, and higher frequency response that the potentiometer type, the Vibrotron gage was sometimes used. Its accuracy stemed from elimination of two analog conversion steps in the over-all data system. Figure 7 shows some typical pressure gages.

For measurement of very low pressures, such as

are encountered at high altitude, the Alphatron was used. Figure 8 shows the gage and associated signal conditioning equipment. Ionization of air molecules was accomplished by use of a radioactive material (Tritium). Positive and negative terminals collected the ions and caused a current flow through a connecting resistance. By selection of resistance, various ranges were covered. Resistance switching hence, range switching, was accomplished automatically. It should be noted that this device measured the number of ions per unit volume, rather than pressure directly.

Calibration of the potentiometer type gages was accomplished by comparison to a calibrated series of weights. Figure 9 is a typical calibration set-up, using a dead weight tester.

Temperatures were measured by means of thermocouples, thermopiles, resistance thermometers and thermistors. The output of thermocouple and thermopile type was automatically compensated for the reference junction temperature by an AMR-developed device known as the "zone box" (Fig. 10), then was amplified by a D.C. amplifier. Resistance thermometers and thermistors were used as one arm of a deflection bridge. D.C. power at about 6 volts was supplied within the signal conditioning equipment. The output from the bridge was amplified by a D.C. amplifier.

Calibration of temperature gages was accomplished with temperature controlled silicone oil baths (Fig. 11).

While there has been occasional use of velocity pickups and strain gage accelerometers for vibration measurements, the JUPITER program relied on the crystal or piezoelectric type for the most part. Vibration transducers are shown in Figure 12. By means of high impedance amplifiers or cathode followers, the frequency response could be as low as 2 c.p.s., using the crystal type. Objectionable line pickup by this high impedance device was avoided by having the amplifier built into the pickup, as is seen in Figure 12. The crystal type had the advantage of higher frequency response over the strain gage type. The velocity pickup's output required subjection to a more complex analysis after data acquisition in most cases in order to yield the information sought.

For pitch and yaw lateral acceleration measurements, potentiometer-type accelerometers were used in both the nose cone and the booster. However, emphasis in the JUPITER was on electromagnetically restrained units, such as those manufactured by Donner. Accelerometers are pictured in Figure 13. These were used because they are capable of the desired high accuracy of 0.1, which was especially useful for studies of longitudinal acceleration at the time of liftoff or cutoff. Their operation was as follows: acceleration acted on a mass connected to a coil, which moved within the field of a permanent magnet. Motion of the mass was detected by a capacitance-changing vane. An electronic circuit dictated a current into the coil so that the motion was restrained. Output voltage was picked up across a resistor in series with the coil. Range switching was performed during flight to take full advantage of the instrument's accuracy. Calibration was accomplished on a centrifuge or a tilt table.

All flow measurements in the JUPITER program were made with turbine-type meters. The passage of a magnet in the turbine past a pickup coil was telemetered so that a frequency proportional to volume flow rate was recorded (Fig. 14). Accuracy was important because we wanted to study details of power plant performance--which would have been impossible without accurate data. By knowing details of performance it was even possible to study aerodynamic drag during flight. As data on mass flow were desired, conversion from volumetric flow was made by reference to density, which in turn was corrected by a temperature measurement. Calibration was made with water. A conversion factor was determined by experiment to yield correct data for liquid oxygen or fuel.

In order to verify information from flow measurements and to accurately determine the amount of residual propellants after cut-off, liquid level measurements were made. Three methods were used to accomplish this. One involved the telemetering of a relay-derived signal each time the liquid gas interface passed a discrete point. The discrete point was determined by location of a piezoelectric probe which vibrated except when damped by liquid. A number of such probes were normally included in the tank. Another method made use of the change in capacitance of a 30-in long concentric cylindrical capacitor (Fig. 15). A servo system then balanced a capacitance bridge. The output was derived from potentiometers on the servo motor shaft. Accuracy of 0.2 inch out of the 30-inch length was possible. This probe was generally located near the bottom of the tank to study residuals. The third method involved a differential pressure measurement between the top and bottom of the tank.

Sloshing of propellant liquids was measured by means of a differential pressure measurement between the two sides of the tanks.

Three types of angle of attack meters were used for R&D measurements. In most cases a potentiometer output was used, although A.C. pickoff systems were employed when the meter was an active part of the missile control system. The three types are shown in Figure 16. They are, from left to right: a differential pressure probe type, a weathervane type, and a wedge vane type. The weathervane type could be used only when an extension boom is permitted on the missile's nose tip. In that location there was less complication of the air stream by the presence of the missile body. However, a boom was usually not practical, so that "local" meters were required. These protrude from the sides of the missile. The probe type had two slots in the protruding shaft at an angle to each other. When the angle of attack was such that the pressure produced in those slots was not equal, the differential pressure produced caused the shaft to rotate until it was equal. The amount of rotation was read out as the position of the slider in a potentiometer.

In speaking of rate gyros, I refer to those used solely for measuring purposes -- not for missile control. Rate gyros are shown in Figure 17. They were used to measure angular velocity in the three axes: pitch, yaw, and roll. The magnitude of rates measured (±10 degrees per second) were small compared to aircraft. While some with D.C. motors have been used, brush wear led us to use A.C. synchronous motors. A.C. pick-offs were generally used to measure the rate of motion. The gyros were spring restrained. The units were calibrated on a rate table.

Error signals from the guidance and control equipment were monitored. Most of this equipment operated from a precision 400 c.p.s. supply. The error signals appeared as amplitude-modulation on 400 cycles carriers. This was converted to a variable amplitude D.C. voltage by the carrier amplifier module or adapter previously mentioned. These error signals were from servo loops which perform such functions as gyro restraint, position actuators, etc.

Various measurements of current, voltage, and frequency were made on the missile's power supplies. For example, one project had a D.C. motor to spin the upper stages. The current drawn by this motor, which at times went up to about 300 amperes, was monitored by a so-called "D.C. transformer". This device consisted of a magnetic amplifier with the equivalent of a one-turn control winding. This winding was actually the lead wire to the motor, and was simply passed through the toroidal transformer core. Thus the disadvantages of shunts and broken connections were avoided. Adapters for current, voltage, and frequency are shown in Figure 18.

One of the power supplies was a 60-volt D.C. supply used in circuits such as feedback potentiometers so that its value was critical. It was monitored in such a way that full scale output of the measurement covered the range of 55 to 65 volts. This was accomplished by means of a voltage divider and a zener diode regulated power supply which cancelled out part of the voltage. A zener diode accuracy of 0.2% yielded a measuring accuracy of 1% for that particular range.

The precision timing system of the guidance system was based on the frequency of a 400 c.p.s. inverter. This was because synchronous motors are used in gyros, etc. In most cases, accuracies between ±0.002 and ±0.015 of rated frequency were regarded as necessary*. In measuring this frequency, the 400 c.p.s. signal was multiplied up to 40,000 c.p.s. and compared to a 40,000 c.p.s. signal from a crystal oscillator. The difference frequency was chosen so that a deviation of ±0.25 c.p.s. in the inverter frequency corresponded to a beat frequency change of ±30 c.p.s., so that analysis on the ground could determine the actual inverter frequency precisely. Crystal oscillators were obtained stable to 5 parts per million over a few hours time. The frequency was checked just before lift-off of the missile.

Sound intensity was of interest for several reasons. High sound levels could have damaging influence on electronic equipment or even cause fatigue in parts of the missile structure. Any plans to carry man will also be influenced by this factor. In large vehicles there is even the possibility of damage to surrounding equipment, people, or buildings. Measurement was made with an ordinary microphone, an Altec-Lansing capacitor-type unit being used in one project. Equipment for measurement of a sound intensity during flight is shown in Figure 19. Equipment for calibration and post-flight analysis of this was available to our laboratory.

Signals from guidance and control equipment, such as gyro positions, actuator positions, signals from guidance computers indicating velocity, displacement, etc., were also monitored.

A new area undergoing study is that of analysis of vibration and sound intensity data within the missile during flight before transmission to the ground. This is of interest because the information bandwidths desired are beyond the telemetering capability. At present vibration is studied only up to 2,000 c.p.s., while there is interest much higher. As analysis will be performed later anyway, it is possible to get around the bandwidth problem by in-flight treatment. To accomplish this we use a small black box. Vibration (or sound intensity) data go into it. Two signals come out: (1) data in terms of amplitude of the data versus frequency, and (2) the amplitude distribution of the data on a statistical basis. While this will not go down to frequencies lower than 50 c.p.s., and transients occurring in less than 4 seconds are missed, a great deal of useful information is made available to aid in design of structures, equipment, etc.

In addition to the multitude of measurements made as part of the R&D phase of missile development, there are often a few special measurements of a more "pure research" nature. Two notable examples which have been widely publicized are measurements on biological passengers and satellite vehicles. Figure 20 shows the signal conditioning equipment used with Monkey Able on her trip into space last summer. A similar but smaller unit accompanied Baker in the same flight, and accompanied Old Reliable in December 1958. This equipment included several circuits of very advanced type which permitted telemetering of such parameters as electrocardiograms, heart sounds, respiratory rate, and body temperature.

Another broad area in which we are engaged is that of scientific satellite instrumentation. The JUNO II consists of a modified JUPITER, with 3 upper stages of scaled down Sergeant rockets. The instrumentation for several NASA missions of the JUNO II is partially the responsibility of our group.

After numerous flights of JUPITER there have

* "Regulation of Airborne Rotating Inverters," W. J. Kreider and M. Jasper, Electrical Manufacturing, 64, 3, Sep 1959, page 103.

been very few failures in the measuring system. Early in the JUPITER program several malfunctions of missile components were pinpointed by the measuring system. Thus the measuring system made direct contributions to the development of the missile. In addition, the measuring system is constantly providing new information on various operations of the missile, which is a valued contribution to our understanding of the advanced missile and space probe projects.

Figure 21 above shows the JUPITER at the conclusion of its 1500 mile flight; a fiery re-entry into the atmosphere.

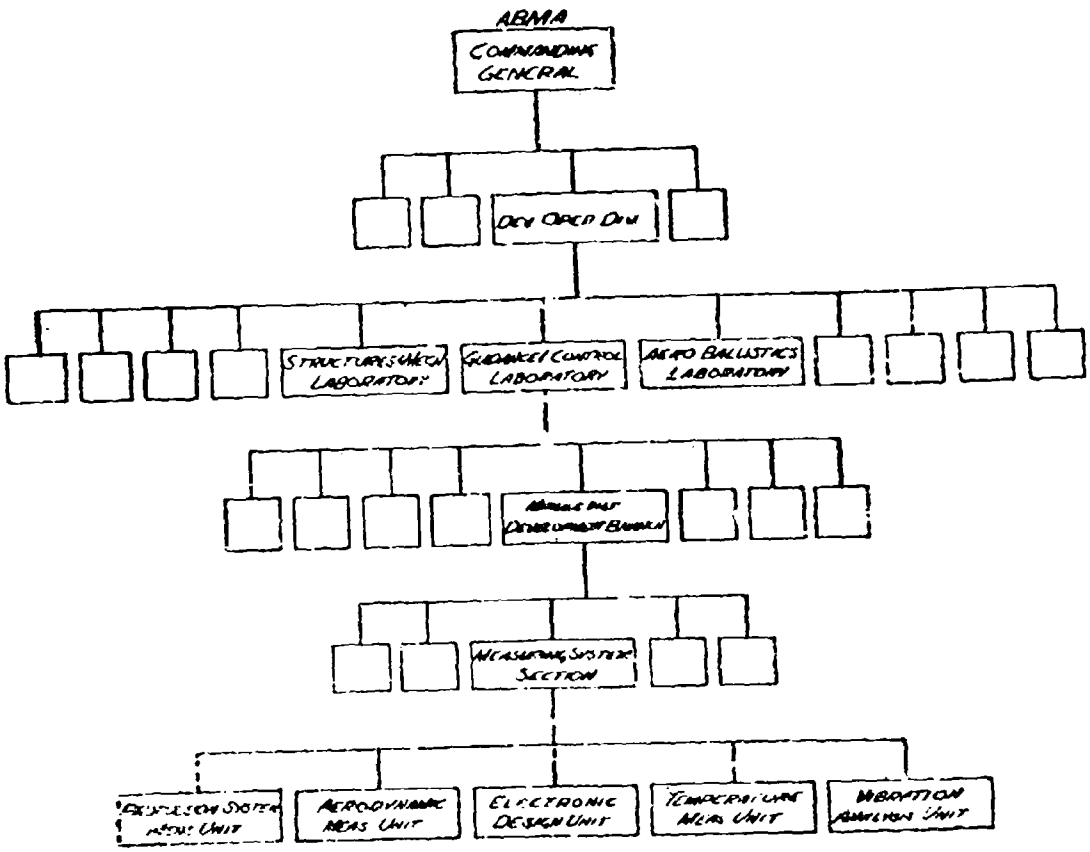


FIGURE 1

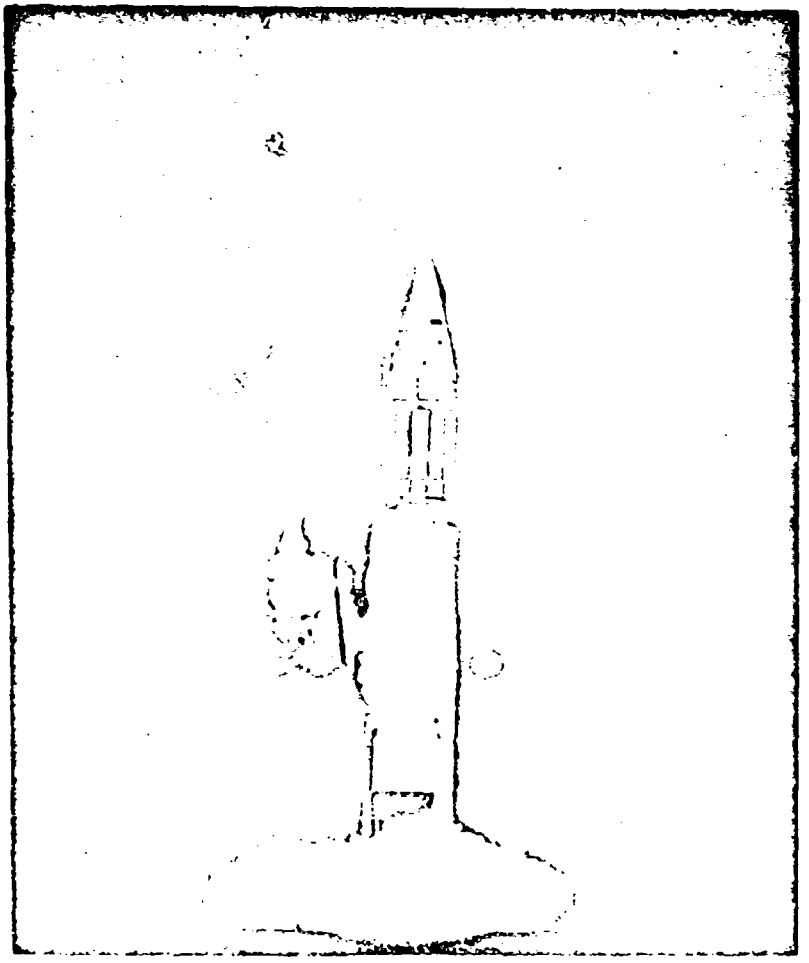


Figure. 2

CONDENSED MEAS PROGRAM, JUPITER

34 PRESSURE MEASUREMENTS
28 TEMPERATURE MEASUREMENTS
10 ACCELERATION MEASUREMENTS
9 VIBRATION MEASUREMENTS
6 ANGULAR VELOCITY
3 FLOW RATE MEASUREMENTS
3 RATE OF SEPARATION MEASUREMENTS
1 RPM MEASUREMENT
1 FREQUENCY OF INVERTER

Figure. 3

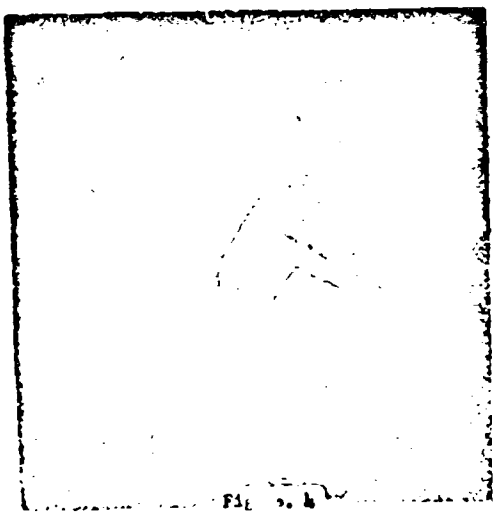


Figure. 4

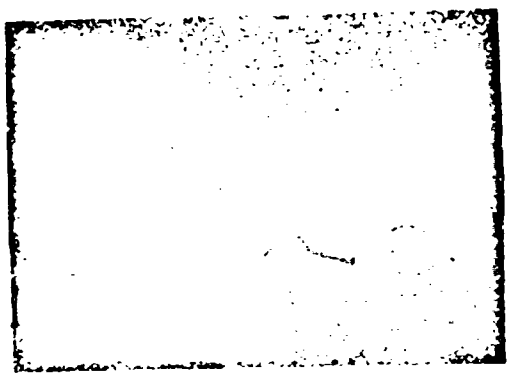


Figure. 7

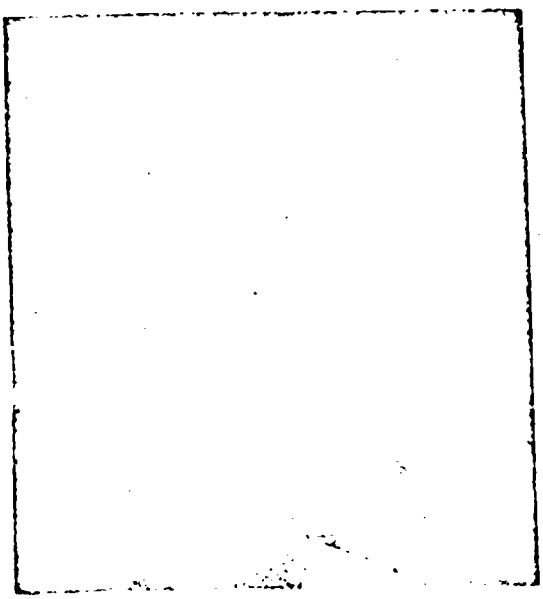


Figure. 5

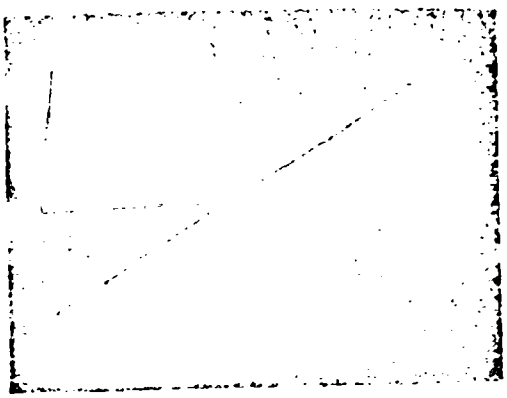


Figure. 8

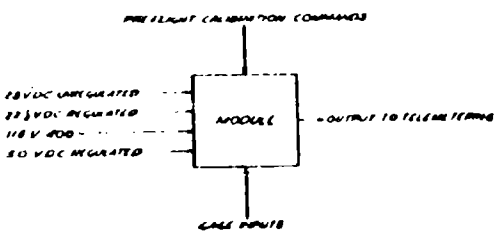


Figure. 6

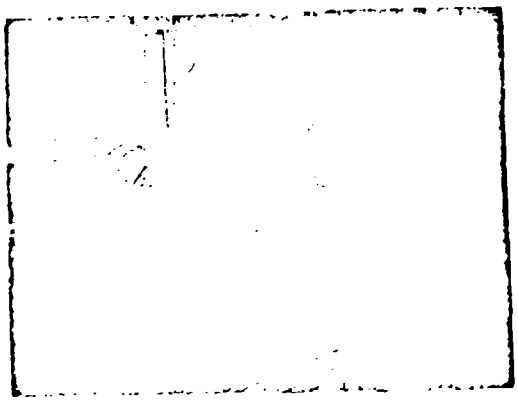
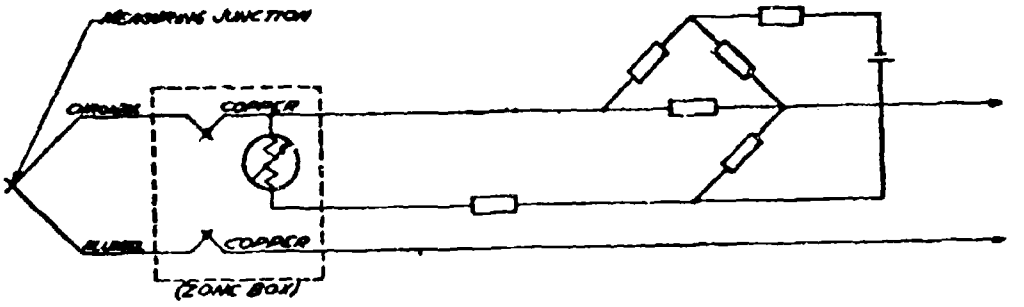


Figure. 9



ZONE BOX CIRCUIT

Figure. 10

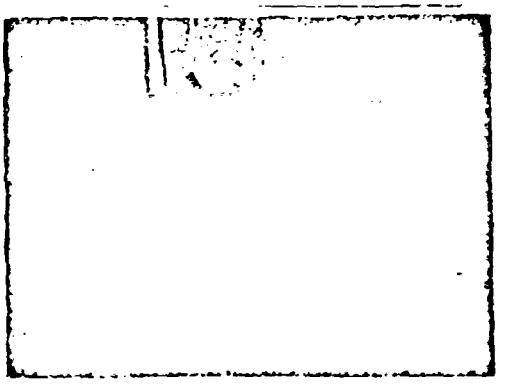


Figure. 11

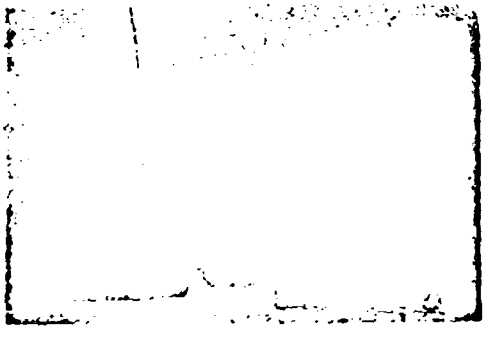


Figure. 13

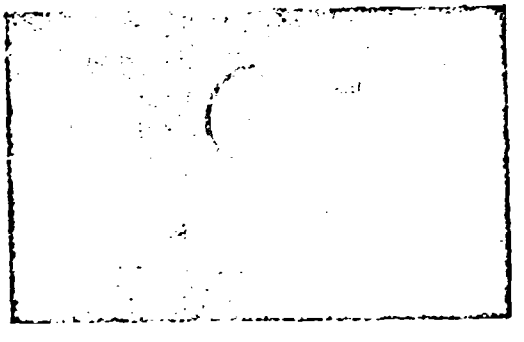


Figure. 12

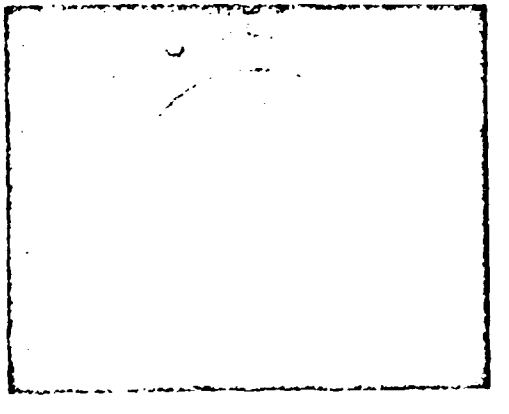


Figure. 14

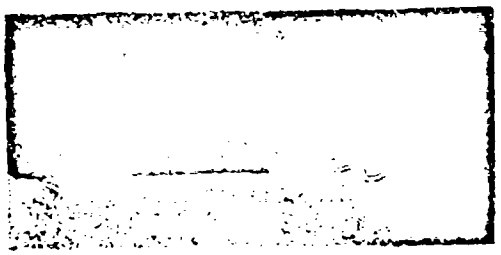


Figure. 15



Figure. 18



Figure. 16

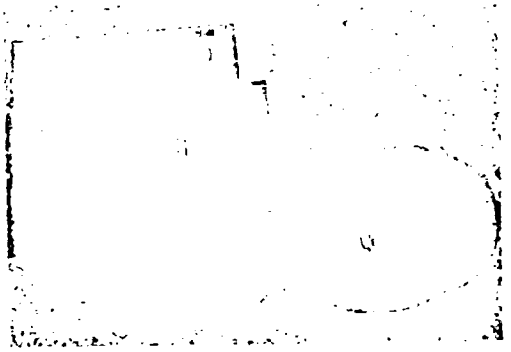


Figure. 19



Figure. 17

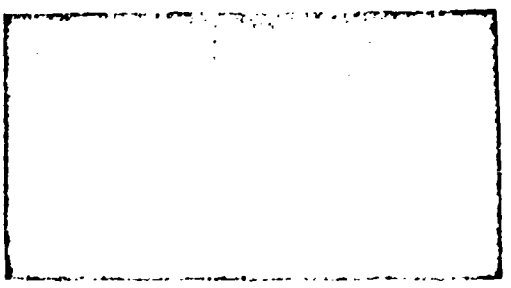


Figure. 20



Figure. 21

LIST OF CAPTIONS

Figure. 1 - Organisational Chart, ABMA

Figure. 2 - The JUPITER Missle

Figure. 3 - Condensed Measuring Program, JUPITER

Figure. 4 - Adapters Used on Early JUPITER

Figure. 5 - The Module System

Figure. 6 - Block Diagram of the Module System

Figure. 7 - Various Pressure Transducers

Figure. 8 - Alphatron Gauge and Adapter

Figure. 9 - Calibration of Pressure Transducers

Figure. 10 - Zone Box Circuit

Figure.11 - Temperature Gauge Calibration Facili Facility

Figure.12 - Vibration Transducers

Figure.13 - Accelerometers

Figure.14 - Flow Meters

Figure.15 - Liquid Level Gauge and Adapter

Figure.16 - Angle of Attack Meters

Figure.17 - Rate Gyros

Figure.18 - Adapters for Current, Voltage and Frequency Measurements

Figure.19 - Sound Intensity Measuring Equipment

Figure.20 - Signal Conditioning Equipment for Monkey Able

Figure.21 - JUPITER Reentry

0.1% ACCURACY VARIABLE SPEED CONTROL SYSTEM

By: Louis Goodman, Defense Products Division, Fairchild Camera and Instrument Corporation

INTRODUCTION

This afternoon I intend to describe two DC to AC Converters. One is used to power the motor in the Fairchild developed KS-25 Hi Acuity Aerial Photo Reconnaissance Camera. The second and more accurate system, now in the final developmental stage, will be used in the Fairchild KS-50 Camera.

In the above two camera systems, as in most reconnaissance cameras, the image is moving while the exposure is being made. To get the ultimate resolution capabilities of the lens film combination it is necessary to move the film at such a rate that the relative motion of the image is zero during exposure. This film motion is called Image Motion Compensation. As can be readily perceived, this motion must be of correct average velocity, and must be free from flutter or instantaneous velocity variations. Certain mechanical errors inherent in all camera designs limit the instantaneous velocity errors to some finite value. The accuracy necessary for the motor drive in such a system can be statistically determined based on the desired resolution and the probable errors inherent in the mechanical design. For the above high resolution cameras the tolerable error in absolute velocity is extremely small. Normal servo methods proved inadequate to the task and a new stable accurate drive system had to be devised to permit camera performance to meet desired specifications.

Figure 1 shows the requirements of the FCIC KS-25. To obtain a resolution of 45 lines per millimeter the motor speed must be held to ±21RPM over a range of about 3000 to 15,000 RPM and since the motor is truly synchronous this means that the motor drive power frequency must be held to ±.7 cps over a range of about 100 cps to 500 cps.

Figure 2 shows the requirements of the FCIC KS-50 camera. Here to obtain the required resolution of 80 lines per millimeter, the motor speed must be held to ±12 RPM over a speed range of about 2000 to 12000 RPM corresponding to a frequency variation of ±.4 cps over a range of about 100 to 400 cps.

The classical approach to DAC Drive is to use AC or DC Tachometers as reference elements in a closed loop servo system. However, the torque variations during a camera cycle are of sufficient magnitude to cause intolerable speed fluctuations. Temperature, and linearity errors are also difficult to control. The input command is a DC signal proportional to the velocity of the aircraft, divided by the altitude (V/H). The AC Tachometer system would make necessary the conversion of this DC V/H signal into proportional AC. This circuitry is complicated and results could be inexact.

A study was made of the constancy of speed of hysteresis synchronous motors under constant frequency conditions, and at the various contemplated speeds. It was found that under full simulated load conditions the observed speed fluctuations were well within the allotted motor tolerance error of ±.15%. The characteristics of this motor made an open loop design very attractive.

The developments in the solid state field within the last few years have made open loop control practical for high accuracy systems, when precise components, are available.

However, the necessary, precise, supporting components were not available either in the required size, or accuracy. Fairchild, therefore, had to develop all the components used in both the KS-25 and KS-50 DAC Drive System, with the exception of the 4 decade potentiometer. (The motor was developed to our specifications).

The KS-25 Drive System

A Description of the Components

This drive system is based on the fact that the rate of flux excursions between saturation end points of square loop magnetic core material can be predicted to within close tolerance, i.e., according to Faraday's Law of Induction, if the flux is changed rapidly from one saturation level to the opposite saturation level, the time taken is directly proportional to the driving voltage: $Vt = -2N \Phi$ max. $X\ 10^{-8}$ so if Φ max. is a constant Vt is a constant.

Where: Vt is the driving voltage (in this case the V/H signal)

t is 1/2f where f is the resulting square wave frequency

N is one half the total primary turns

$\Phi = B_M \times A$ where B_M is the saturation flux and A is the cross sectional area of the core.

The entire design was therefore, based on preserving the constancy of this driving voltage and the core saturation flux.

Figure 3 shows in block diagram form the system that resulted.

The heart of this system is the two phase flux type of multivibrator. By an exacting effort, it was possible to make this relatively simple circuit give the desired performance.

KS-25 Camera System
Data Sheet

Type.....Hi Acuity

Picture Size.....4-1/2"x 4-1/2"

Image Motion Compensation.....4 to 1.7"/sec.

Lens.....24" Focal Length - f/4, High Resolution, Spima Lens
80 Lines/MM AWAP - low contrast (2:1 - Ratio in Surface
Brightness) Automatic Focusing - Lens Position Accuracy
± .001"

Shutter.....Focal Plane Type

Shutter Speed.....1/25 to 1/1000

Resolution.....45 Lines/MM at 1/25 second Low Contrast

Max DMC Speed Deviation.....To obtain required resolution ±22.5 RPM (±.75 cps)

Size.....11" deep x 14" deep x 40-1/2" long

Estimated Total Weight Loaded.....150 lbs.

Operational Temperature.....65°F to 120°F

FIGURE 1

A transistorized operational amplifier module was developed and is used both for the high accuracy power supply and the isolation amplifier. A .1 Decade potentiometer supplies the V/H voltage, through the isolation amplifier to the flux oscillator. Its output is a square wave. The linear integrator converts the square waves into triangles which finally appear at the motor terminals after going through a buffer amplifier, power amplifier, and coupling network.

The operational amplifiers modules are solid state chopper stabilized; the input is differential, and the output single ended. The band pass is narrow to minimize noise. Open loop gain is 1.5 $x10^6$. Equivalent input noise level is about 1 millivolt. At 60°C it is capable of supplying a maximum power output of 1 watt at ±22 volts. The input is differential, permitting the use of low impedance feedback elements which do not load the signal input. In the power supply unit the amplifier gain is adjusted by feedback to multiply the stabilized zener referenced voltage. It develops 21.0 volts across the V/H potentiometer.

A similar operational module used as a unity gain follower amplifier isolates the V/H potentiometer and supplies the V/H voltage at a low impedance level to the 2 phase flux multivibrator. The small plug-in package containing the two amplifiers is shown in Figure 4.

The two phase flux type of free running multivibrator uses transistors in their non linear region to switch the flux in square loop magnetic cores. The circuit is regenerative, and develops output square waves whose frequency is determined by the DC driving voltage. The driving multivibrator is tapped supplying driving current which forces another multivibrator having identical characteristics into oscillation. This slaved oscillator develops an output voltage in exact quadrature with the forcing multivibrator. Correct performance is obtained by exact windings and careful matching and biasing.

The output voltages of the multivibrators have a rising characteristic of 6 db per octave. The strong harmonic content of these square waves would cause excessive heating of the output transistors and motor. A linear integrator with a drop off of 6 db per octave produces a triangler wave form. This voltage is constant. The two buffer amplifiers are used to prevent loading and drive the class "B" output amplifier. This amplifier is coupled to the motor fields through a dual variac assembly servo drive in accordance with the V/H signal. These variacs make the motor load appear constant to the low resistance transistors, and permit the motor to develop a constant torque. A second regulated power supply converts the unregulated dc input power into regulated 1600 cps square waves and regulated ± 28 volts. The square waves are the carrier for the operational amplifiers. The regulated ± 28 volts power the operational amplifiers, buffer amplifiers and the Zener reference.

CONCLUSION

The precision with which the output frequency follows the driving V/H voltage is shown in figure 5. These curves show the linearity error at various temperatures. The maximum uncompensated error at room temperature is .3% of 500 cps., or 1.5 cps., and occurs at about 250 cps. By error voltage injection into the isolation amplifier feedback loop, it is possible to reduce this error to less

than .08%. The effect of this compensation is shown in the corrected linearity curve at 74°F.

The effect of temperature is to increase the frequency .05% /°C. This is a linear error. The orthonal core changes its saturation flux level .07% /°C. This is compensated by an opposite IR error of .02% /°C in the transistors and windings.

Temperature compensation is obtained by sensing the core temperature with a non inductive resistance thermometer on the core. This resistance change is in the feedback loop of the high accuracy power supply and varies the voltage supplied to the V/H pot. The flux oscillator is therefore supplied with a temperature sensitive voltage in such a direction as to offset the change in frequency. The maximum corrected error observed is less than ±.1% or ±1/2 cps. Stability and repeatability is better than .04% at any one temperature. Figure 6 and 7 shows the motor velocity variation at 3532 RPM, and at 15,162 RPM. These curves were obtained under operating conditions by observing the change in velocity of a tone wheel. The oscilloscope calibration is 1 vertical centimeter equal to a velocity deviation of 1% of the maximum frequency.

A summary of these results is tabulated below:

Motor Speed RPM	Motor velocity change during exposure -%	Effect on Image motion during 1/25 second exposure inches
3532	± .1 (3.5 RPM)	.00001
8830	± .15 (± 13 RPM)	.00004
15,162	± .5 (±23 RPM)	.00007

KS-50 Camera System
Data Sheet

Type.....Panoramic	Twin-scan synchronized with single mechanism. Interchangeable cassettes.
Coverage....150°	2 lenses each covering 75° from vertical with 5° vertical overlap. Forward overlap 55%.
Film.....9-1/2"	5000 ft. thin base (SC-1213) 2500 ft. standard base (Plus X)
Weight (Est.)	Camera - 400 lbs. Mount - 150 lbs. Mount electronics - 30 lbs. Control Box - 10 lbs.
Overall camera size	Approximately 40" cube.
Operating Range	20,000 to 100,000 ft. altitude 300 to 2500 knots ground speed
Operating Temperature	0°F to 120°F
Cycling Rate Range	2 sec/cy to 20 sec/cy.
V/H Range	.05 to .005 (10:1)
Shutter Speed Range	AEC provides 1/10 to 1/2400 (.026 to .625 slit width).
IMC Rate	2"/sec. to 2"/sec.
Scan Motor Speed	3750 to 12,000 RPM
Lens	24" f/5.6
Resolution	80 lines/mm at 1/50 sec shutter speed (bench test)
Maximum IMC Speed Deviation	To obtain required resolution ±12 RPM (± .4 cps)
Mount	Torquer stabilized Steadiness - 2 seconds of arc in 1/25 second. Verticality - 3 seconds of arc. Degrees of Freedom - ± 8° in pitch, roll and azimuth.

FIGURE 2

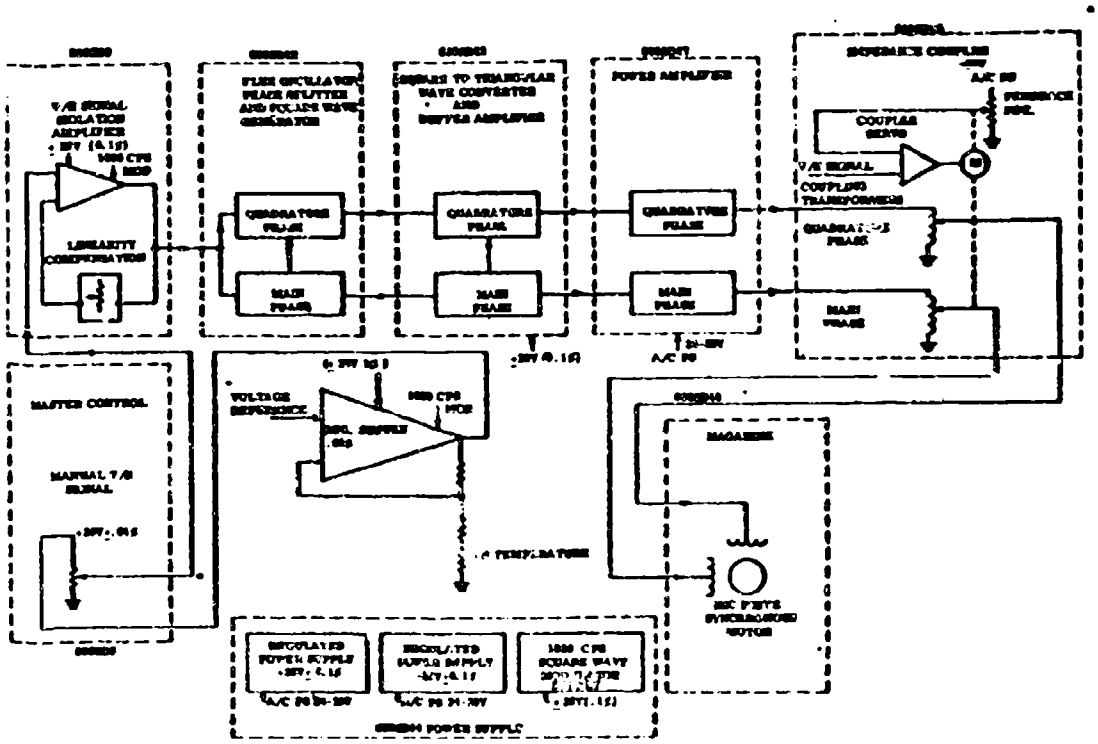


FIGURE 3 - BLOCK DIAGRAM KS-25 DRIVE SYSTEM

This circuit has been breadboarded using newly developed Low RCS Silicon Transistors. Figure 8 shows the results obtained without compensation.

No attempts were made at correction. At room temperature the linearity error is within .1%. The temperature rise curves are nearly linear and are very close to the theoretical core error. Indications are that these transistors supply little or no deterioration to the system performance.

SUMMARY

Summarizing the first part of this talk, Fairchild has developed a high accuracy DC to AC inverter for a DC motor drive for an aerial camera. The circuitry is simple and this drive system easily accomplished a task that previously required a high degree of circuit sophistication.

Part II

KS-50 DMC Drive System

INTRODUCTION

The greater accuracy requirements of the KS-50 DMC drive made necessary a refinement of the simple open loop system. This system was discussed in part I of this talk. In the simple system the generating oscillator is also the frequency reference. In the KS-50 system the linearity and temperature errors of this generating oscillator are corrected by reference to a temperature controlled magnetic square loop core standard. The design of this volt second reference will be discussed at greater length later in this talk.

DISCUSSION

Figure 10 shows in block diagram form the corrected oscillator DMC drive system.

The added components, other than the volt second reference and its drive system are associated with summing the error difference signal between the reference and the DC V/H voltage. An added operational amplifier has been inserted between the isolation amplifier and the sine wave oscillator, to prevent the V/H pot resistance variation from entering into the summation.

Briefly, the constant voltage drive circuit limits the square wave output out of the master flux oscillator (0°) to a constant voltage level, but, at a higher power level. Its output switches a pair of low saturation resistance transistors. This switch action rapidly changes the flux level of the volt second reactor from one saturation level to the opposite saturation level, then resets. The entire volt second reference circuit, and also, the Zener reference diode, are in a constant temperature oven. The constant area pulses appearing in the secondary are rectified. The width of these pulses is determined by Faraday's law of induction as mentioned in Part I. Their amplitude is determined by the fixed DC voltage to the collector, and their periodicity by the driving frequency. The negative pulses are compared with the positive V/H voltage from the isolation amplifier, utilizing a Varley loop. The error difference is fed to an integrator connected operational amplifier. A resistance network at the input of the summing amplifier compares this integrator averaged DC error voltage to the V/H voltage. The difference appears at the output of the summing amplifier. This voltage is corrected for both linearity and temperature variations of the flux oscillator.

The curves in figure 11, show the uncorrected linearity error. Also shown in the same figure is the corrected linearity error using the same linearity correction circuit as is discussed in Part I. The oven temperature is held to about 1/20° C at the core of the volt second reactor. The total error for both linearity and temperature is about .05% of 400 cps or about .20 cycles per second. The maximum error allocated to the oscillator is ±.4 cps. The results are 4 times better than specified. This system is in the final developmental stages. It will be interesting to see how much more accurate it can be made.

The Volt Second Reference

Figure 11 shows the circuit of the Volt Second Reference. This system is based entirely on the constancy of the volt second area of magnetic material when subject to saturating flux excursions. From the integrated Faraday's law formula that; $Vt = -2 N\phi$ max. we can see that Vt is a constant if ϕ max. is a constant. ϕ max is not a constant since there is a changing slope to the saturation curve which can vary as much as 4 to 1. As the induction increases the slope decreases. Temperature error must also be considered.

"V" is a summation of Vcc minus all the series IR drop. These IR drops are found in all ohmic elements such as winding resistances, transistor resistance, primary and secondary resistance and diode resistances. Other sources of error which can be lumped in with the IR drops, are the hysteresis and eddy current losses. These are non linear and increase with frequency.

The volt second area reference has been designed to minimize the aforementioned errors. The switching transistors are driven by a constant voltage source yielding a constant core excitation. The ampere turns are high and constant so that ϕ max. is set in the region of low permeability.

Care has been taken to minimize IR variations. The transistors are silicon type and have low saturation resistance. The rectifying diode and transistors as well as the core are in a temperature controlled oven.

Since the pulse width is constant and the sides are steep, eddy current losses tend to be constant.

The accuracy requirements of this drive are such that the errors usually considered of secondary importance in most systems must be considered.

CONCLUSIONS

Fairchild has developed two accurate motor drive systems for image motion compensation in aerial cameras. These systems are dependent on the concept of rapidly switching saturable cores to precisely determined end parameters. The data indicates that a high order of accuracy is obtained from the combination of transistors and magnetic circuits. These elements have provided simple solutions to formidable circuit problems while increasing circuit reliability. This method permits the generation of a band of frequencies to an accuracy previously possible only at a single frequency utilizing tuning fork techniques.

ACKNOWLEDGMENTS

The author appreciates the cooperation of Ralph Lindberg and Joseph Kaplan who worked on various phases of this project, of Melville Hartmann, who was Sr. Project Engineer during the initial stages of the KS-25 development, of Paul Knight who corrected the original manuscript, also Theodore Berthold and Harry Hastings, who were the project chiefs of the KS-25 and the KS-50 Aerial Cameras. This work was performed under contracts AF33(600)35838 and AF33(600)38241.

REFERENCES

1. G. H. Royer, "A switching transistor DC to AC converter having an output frequency proportional to the DC input voltage AIEE Transactions Part 1, Vol. 74 P 322-325."

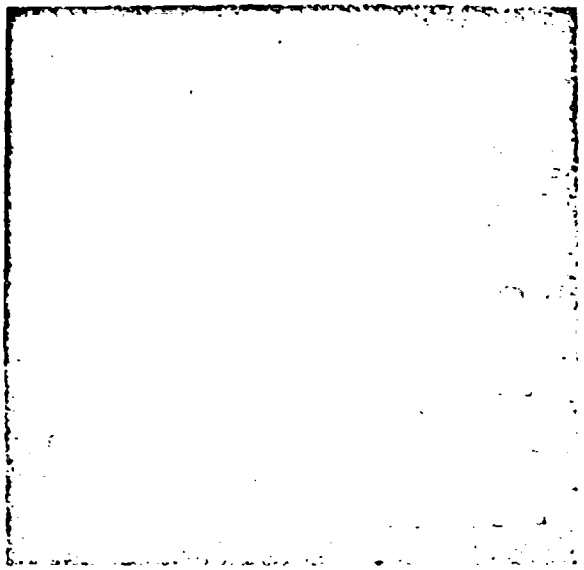
2. J. L. Lawrence, "Magnetic Saturable Core Timing Device" Communications and Electronics July 1958 P 393-396.

3. R. L. Van Allen, "Time Interval Generation with Transistor Switches and Magnetic Cores." Private paper of Mr. Van Allen.

4. R. A. Anderson & H. B. Schulthesis, "New Voltage to Digital Conversion Technique - Electro-mechanical Design Nov. 1959."

5. W. H. Card, "Transistor - Oscillator Induction Motor Drive - AIEE Transactions Paper #58-898."

6. Barry Miller, "High Acuity Photo Reconnaissance Studied", Aviation Week Jan. 11, 1960.



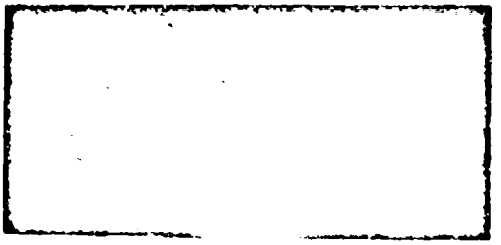


FIGURE 6



FIGURE 7

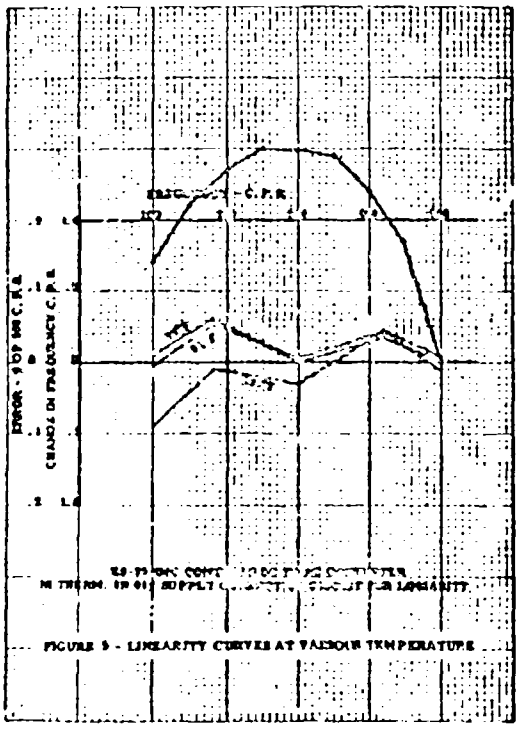


FIGURE 5 - LINEARITY CURVES AT VARIOUS TEMPERATURE

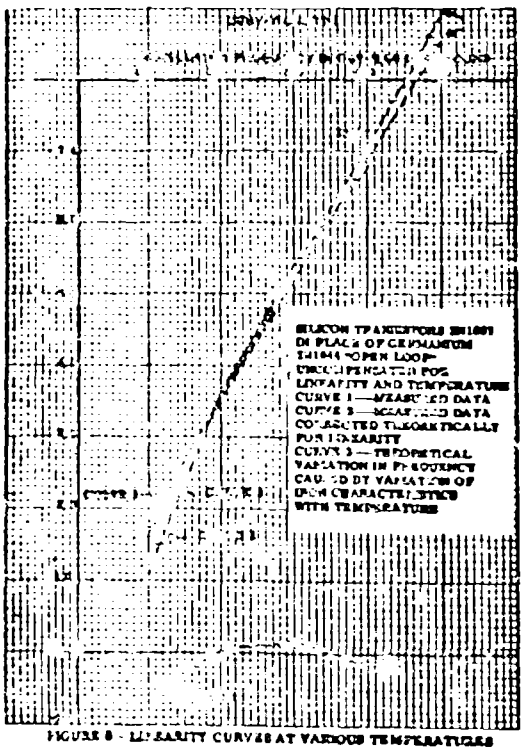
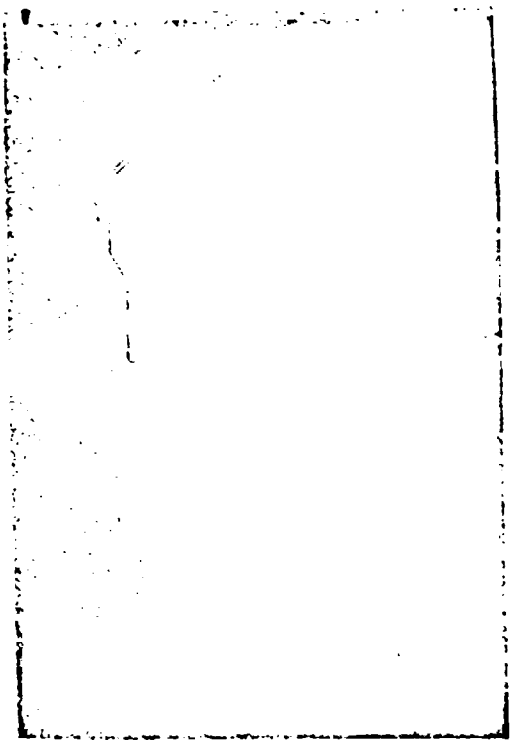


FIGURE 8 - LINEARITY CURVES AT VARIOUS TEMPERATURES



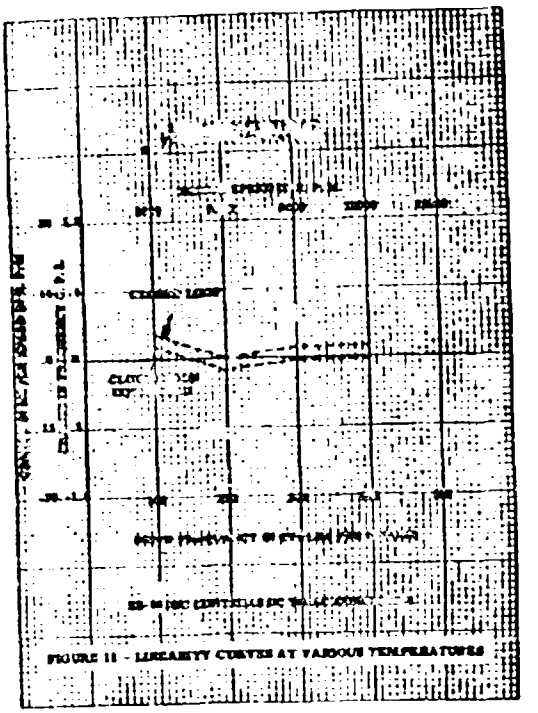
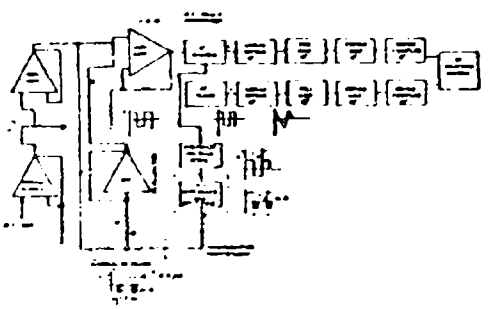
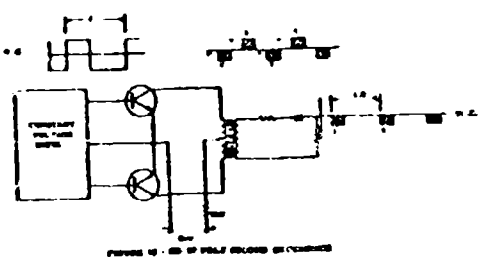


FIGURE 11 - LINEARITY CURVES AT VARIOUS TEMPERATURES





THE IMPLEMENTATION OF THE INTEGRATED MAPPING SYSTEM

By: Arthur Fein, Defense Products Division, Fairchild Camera & Instrument Corporation

Virtually everyone connected with the operational phases of the photogrammetric processes as applied to cartography has had in his own plans the application of an automatic process of matching images and thus obtaining automatic stereo plotting. Steps are being taken which will bring this into reality.

Camera and supporting equipments are reaching higher performance levels. As a source of error the camera is virtually no longer a problem. Lens distortions can be maintained at tolerably negligent levels, film and processing can be controlled. Instrumentation for the purpose herein described must, therefore, not destroy the accuracy contained in the diapositives from which it works.

A modification of the approach to the production of accurate orthophotoscopes, the development of which was explained by Russell K. Bean in 1955, provides the basic used in this instrumentation for the orthophotoscopic record.

The concept of contour delineation through the act of profiling, as revealed in the December 1957 issue of Photogrammetric Engineering by Spooner and his associates at Army Map Service, are brought to fruition in this instrumentation. This paper is (in part) a progress report based on the implementation of the principles proposed by the Army Map Service at that time. The merits of profile scanning instead of contour tracing for the purpose of hypsometry were convincingly treated by Messrs. Spooner, Doonl and Misulia. Benefits accruing to the operator, as regards time reduction and operator training are evidently significant.

The three products of the Integrated Mapping System as conceived by the Army Map Service and currently being built by the Fairchild Camera and Instrument Corporation are:

1. Plotted Contour Information

2. Orthographically Positioned Photographic Detail, and

3. Stored Profile Data

The first product, of primary interest to all photogrammetrists in the cartographic field, must be contour information in a suitable form for the compilation of elevation data measurements to standard map accuracies.

The second, commonly known as the orthophotograph, can be used as a source of compilation of all planimetry in the operational phase of cartographic work. The imagery converted from a perspective view of the terrain as it appears in the aerial photograph to an orthographic view must be sharp and contain enough detail for easy recognition and interpretation and must meet the required standards for horizontal accuracy. Accurate orthographic proportions are not present in any case where the photographed terrain does not lie in one plane.

Tests made for comparison of the effects on the final model of profile scanning, compared to conventional techniques of "chasing a contour" throughout the model, have shown the completeness of the profile technique to be superior. This was evidenced by the absence of isolated hill tops using the conventional methods. The improvement of course, is the consequence of the fact that profiling compels the stereo operator to examine all parts of the model in detail.

The third product, a permanent record of the data used in the hypsometry becomes a ready reference for use in a number of applications, as, for instance: map revisions; guiding the cutting arm of a terrain model cutting machine; and thirdly for data conversions such as vertical relief exaggeration or change of datum by utilizing a high speed electronic computer.

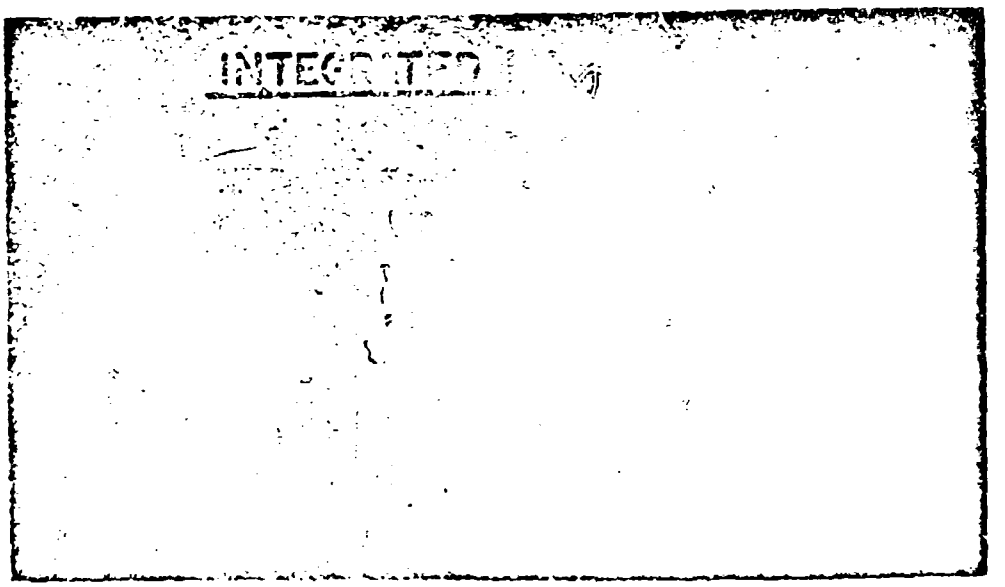
Figure 1 shows the three working components of the system. It does not include the electronic control system, which contains the temporary and permanent tape handlers, nor the operator's control console. The Nistri Stereoplotter, furnished by the government for this project, has been modified by the substitution of more powerful servo motors to drive the X and Y arms of the coordinatograph, and the Z axis assembly.

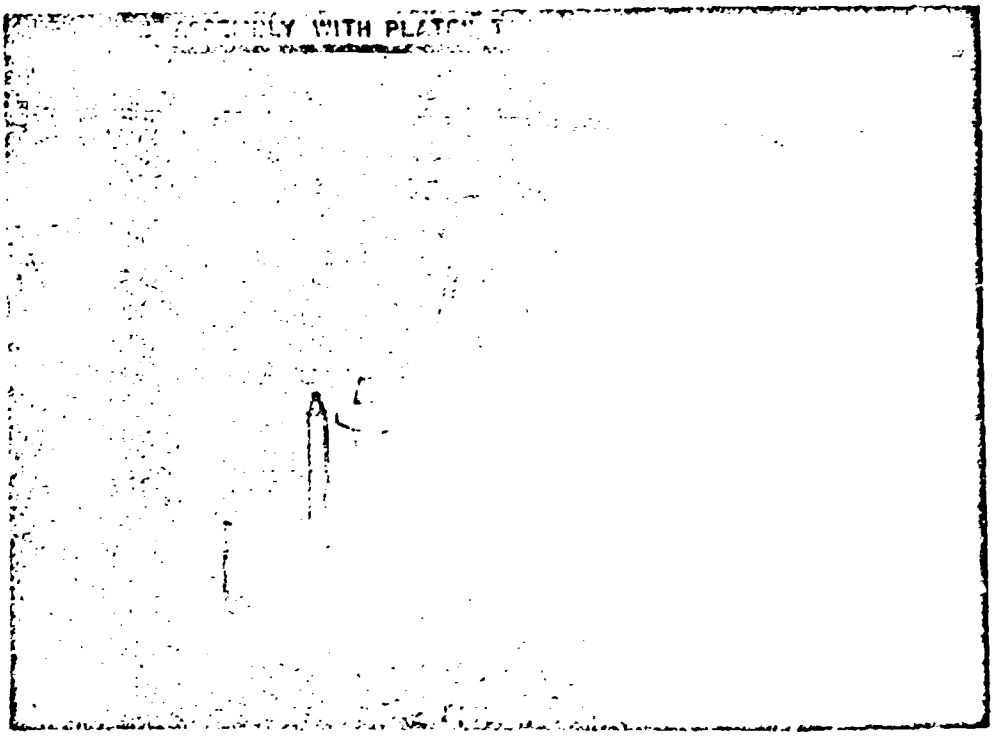
The line drop contour plotting table, shown at the left, contains a similarly modified coordinatograph used to transport the line drop marking assembly.

The orthophotoscope unit, shown at the right, also employs a modified coordinatograph which drives the orthophoto exposure unit. Exposure is obtained from a cathode ray tube operating from a closed circuit television system; the entire working system is enclosed in a light tight housing for working convenience.

Figure 2 shows a close-up view of the platen assembly modifications. The "black box" at the left, adjacent to the circular platen, is the television camera pick-up tube, used to make the orthophoto. Together with the platen, it slides horizontally on a track which is driven vertically by the Z drive system. When the operator scans the profile the platen is in operating position. When the machine is recording the orthophoto, the camera tube is indexed horizontally into position. They both move up and down in the Z direction regardless of which is in operating position.

Figure 3 shows the electronic control system.


INTEGRATED


WITH PLATE

Three magnetic tape handlers can be seen at the left. Two of these are used for the control and recording of X, Y, and Z data on a temporary basis; the third is used to form the permanent tape record of the model area. At the upper right a television monitor screen can be seen. This is a monitor for the closed circuit television system and provides the operator assurance that the orthophoto, being formed in its light-tight enclosure, is receiving its video signal from the camera.

The remainder of the electronic control system consists of the servo control for the coordinatograph arms and Z drive, the digital data handling circuitry for the acquisition, processing and control of the X, Y and Z information, and the several power supplies required by the system.

Figure 4 shows the working system. This shows the tie-in-between the various components. Each of these units, the tape handlers, the stereoplotter, the line drop and orthophoto units, has just been touched on in reviewing the previous slides, and will be referred to again.

Through the act of profiling, the instrumentation produces all of the data necessary to produce topographic maps, master terrain models and allied products.

As just stated, the instrumentation produces as direct outputs, a permanent record of the profile data on magnetic tape, a contour plot in the form of a line drop compilation, and an accurate orthophoto.

Profiling is accomplished with greater speed and accuracy through the assistance of a memory system. A temporary record of the previous profile actually drives the floating mark, requiring the operator to merely provide corrective adjustment. The over-riding adjustment made by the operator continually results in the formation of a new temporary record and may be repeatly played back for review and correction.

The useful outputs, produced automatically by the machine are made only when the operator is satisfied that the temporary record accurately contains the desired profile data.

INTEGRATED MAP SYSTEM CONTROL SYSTEM

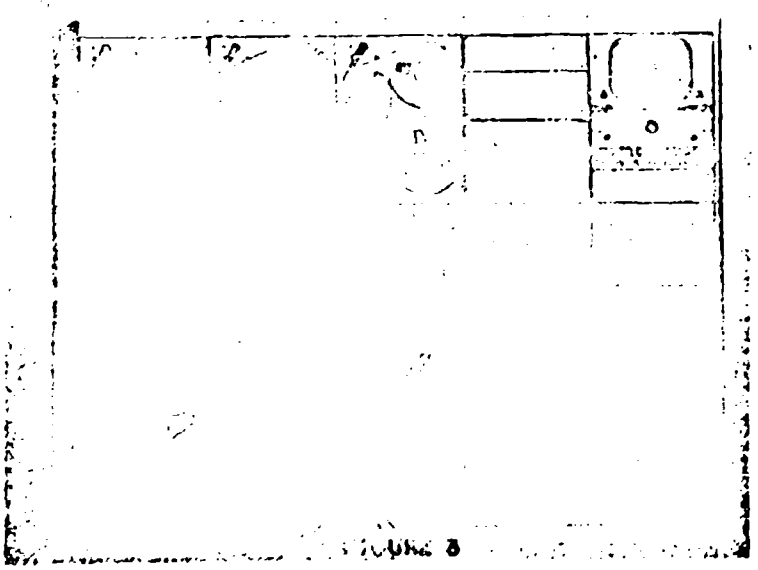


Figure 3

Some of the significant data relative to the performance of the machine will now be presented, before continuing with a further explanation of its operation. The three tables have the same X and Y travel capabilities:

In the X direction:

a) The maximum travel is 800 mm. This pertains to the steroplotter, the line drop head and the orthophoto exposure head.
b) Position is known and recorded to an accuracy of 1/10 mm.
c) The speed of travel is controllable by the operator from 0 to 20 mm per second.

In the Y direction:

a) The maximum travel is 1000 mm.
b) The Y increment value is adjustable in steps of 1/10 mm up to a maximum increment of 12 mm and is recorded to an accuracy better than .025 mm.
c) The Y step-over time is a maximum of 72 seconds for a Y increment of 12 mm. (6/10 seconds per 1/10 mm step)

In the Z direction of the stereoplotter:

a) The maximum travel is 180 mm.
b) The actual Z value is recorded each 1/10 mm of X travel to an accuracy of 1/10 mm.
c) The maximum speed of travel is variable from 0 to 20 mm per second.

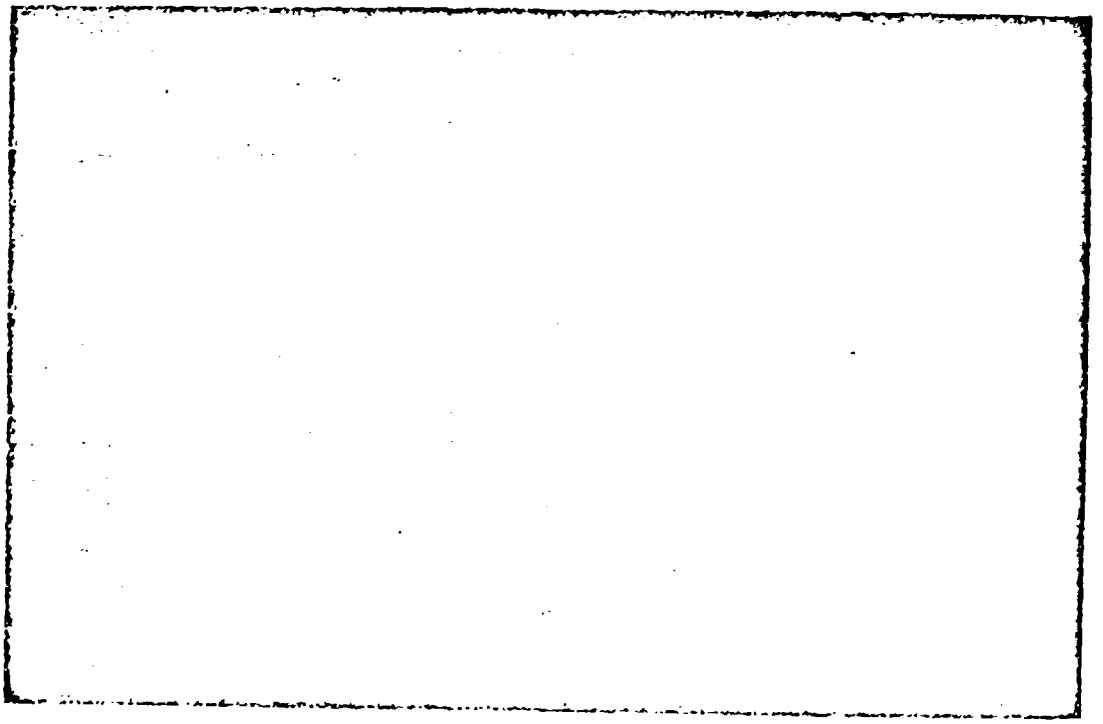
The line drop compilation has the following performance:

a) The frequency with which lines are dropped is adjustable to accommodate 4 contour level increments of 20 and 40 feet, 10 and 20 meters, and 9 model scales ranging from 1 to 3600 to 1 to 10,000. These correspond to Z increments which range from 1 to 4 mm.
b) The accuracy of the dropped line end is 1/10 mm.
c) Line dropping speed is compatible with the scanning speeds in X and Z. Approximately 40 half contours can be recorded per second.

Lastly, the orthophoto system:

a) The width of the exposed strip in the Y direction is adjustable in 1/10 mm steps, up to a maximum width of 12 mm.
b) 6 reduction ratios ranging from 1 to 1, to 1 to 6 are provided.
c) The resolution capability of the orthophoto is 75 percent of that obtained at the scanning plates of the stereoplotter.

Returning to the operation of the machine; profiling is done in the "X" direction, (parallel to the flight direction.) In the operating unit the platen is driven at a selective speed in the "X" direction, (up to 20 mm per second.) While the operator "rides the profile", the machine



will temporarily record the profile. Realizing that there is room for error in the first scan, the operator may at the end of the scan, reverse direction and let the platen be positioned up and down in Z by the magnetic tape record just made. Should descrepancies occur, corrections can be made "on the spot" by an operator's over-ride control, resulting in a new recording of the entire profile.

Once satisfied with the profile, the operator may cause the machine to make the line drop profile record sheet, an orthophotograph, and a permanent recording on magnetic tape in any reasonable machine language for electronic computers. The language which has been selected for this machine is for use in the UNIVAC Model 1103A or 1105 computers.

During the operation on the second or adjacent profile, temporarily storaged information from the previous profile is used to position the platen in Z, modified by the operator's override control to produce the recording of the new data. This means that only a single profile in the model need be done with no previous knowledge of where the platen should be. Once accomplished, the memory aided mode of operation is fully effective. On a single profile, this information is available for refinement in the "X" direction or along the profile. On adjacent profiles, this information is available for the same type of refinement, but in the "Y" direction.

The memory aid profiling is done with an electronic analog to digital converter capable of high resolution and an output of digital information at a rate of about 2000 elevations per second. Since the rate of scanning, as presently called for, is a little less than one inch per second, almost an infinite number of readings of elevations are available to provide smooth operation.

As mentioned above, once the operator is satisfied with a profile, he may produce a contour sheet. This is done with the line dropping mechanism. Throughout the range of elevation (Z motion of the platen), there is provided a series of points corresponding to the selected contour intervals, which, while the profiling operation takes place, will operate the line drop device.

This device will mark continuously from one elevation to the next, with alternate contour intervals left blank. This system has been further advanced from the compilation standpoint so that line identification will be provided by color coding. A compiler will be able to tell at a glance whether the slope is up hill or down hill, in any direction, and in any local area without reference to control points or drainage patterns.

As an example of this system, if the slope is up hill the sequence of collor coding will be:

Red, Blank, Black, Blank, Green, Blank, Red, Blank, Black, etc.
A down hill slope would be just the reverse.

In other words, if the sequence (not counting the blanks) goes from red to black or from black to green, then an up hill slope is indicated. A down hill slope would be indicated by a sequence of black to red or from green to black.

As the second product, which can be made concurrently, the instrumentation being developed by Fairchild will provide an orthophoto of the model area. This is provided by mounting on the platen, a television camera which takes a picture of the local terrain (i.e., in the area of the floating mark). The actual exposure will be of an area of a few thousands of an inch in the "X" direction by the width of scan in the "Y" direction equal to the "Y increment", and will be obtained from the projection of one of the diapositives of the stereo pair, with the red filter removed.

After being amplified by a remote video amplifier, this signal is reproduced on a small Cathode Ray Tube which is mounted on a second electro coordinatograph. The remote CRT is mounted so as to make an exposure of all of the details in the photograph on to a flat sheet of photographic film.. Proper photographic filtering will prevent the second photo of the model from being reproduced, or one light may be turned off, since during this operation no stereoscopic viewing is required.

The CRT reproducer travels only in the "X" direction during a single profile (i.e., no Z motion is included in the making of the orthophotograph.) Thus, a flat photographic exposure plane is used. True orthographic proportion is obtained by controlling the Z motion of the television camera pick-up tube from the tape recording. This technique, as some of you know, has the advantage of not having to raise and lower a large flat photographic sheet during the profiling operation as would be the case if the exposure slit and film were in the actual model space. A second advantage is provided in the Integrated Mapping System concept in that the operator is given a chance to change his profile to a completely satisfactory one before making an exposure on the orthophotograph.

The third product of the Fairchild instrumentation is the provision for making a permanent recording of the model space. This can be done either concurrently with the line drop contour chart and/or the orthophoto; or it may be made separately. Operationally this does not cause any difficulty. It is expected that the three products, contouring, orthophoto and permanent recording can be accomplished with only two scans of each profile (except the initial one).

Thus: Right to left on the initial scan will provide course data.
Left to right will provide a check, or
Left to right - contouring, exposing and recording.

On subsequent scans:

Right to left will provide a check with minor corrections.
Left to right - contouring, exposing and recording.

AN EVALUATION OF TECHNIQUES IN LAND MASS RADAR SIMULATION

George M. Dembo and Charles Colbert
Westgate Laboratory, Inc., Yellow Springs, Ohio

Summary

An objective evaluation of the various land mass radar simulation techniques in use today is useful in determining how a specific requirement may best be met. The capabilities of the ultrasonic, optical analog and two transparency techniques are examined and compared. Specific examples of work done in the various techniques and some of their advantages and limitations are cited.

Map projection gaps and distortions are discussed and a simulator which overcomes this basic problem is briefly described.

Introduction

Is there a unique solution to the problem of land mass radar simulation? The answer is no. There is no universal simulator or simulation technique which is so versatile, so flexible that it can meet all of the complex requirements of today's weapons systems. A look at the development of the land mass simulation art shows us that the various techniques in use today have been developed in response to a wide range of requirements. Because these requirements change with new weapons and new tactics philosophies, it is impossible to develop the universal simulator which meets all existing and future needs.

Unfortunately, as each new technique is developed it claims to be the long-awaited simulation panacea -- one which will answer all requirements. Such is not the case. Each of the techniques in use today has its advantages and its limitations. Each has - and deserves - its place in the sun, depending on the job it must do. Therefore, a close look at the capabilities of each technique is of value in making a selection to meet a specific set of requirements.

Simulation Techniques

What are these simulation techniques? By way of definition, the techniques referred to here are those in use for simulating the land mass radar returns as seen in airborne mapping radars used for navigation, bombing, search and tracking, and as used in the preparation of reference data for "map matching" missile guidance systems. These techniques may be described generally as being one of two basic types: analog or computational. Let us consider the analog systems first. They are so called because they mimic the action of the actual radar by radiating energy (other than microwave) over the surface of a terrain analog in the form of a three dimensional relief map.

There are three analog techniques in use today. The first of these, the ultrasonic technique, has been employed since World War II in simulators and trainers for both the Air Force and Navy. Essentially, this technique simulates land mass radar returns by radiating pulsed ultrasonic energy, using water as a propagation medium, from a crystal transducer located at the simulated antenna position. The radiated energy is concentrated in a narrow beam as in the real radar case. Sweeping out in range and scanning in azimuth as in the real case, the radiated energy strikes the surface of a specially prepared terrain model. Figure 1 illustrates schematically the basic operation of this technique. The re-radiated energy reflected from the terrain model surface is received by the crystal, and is converted to a video signal which is ultimately displayed on the radar indicator.

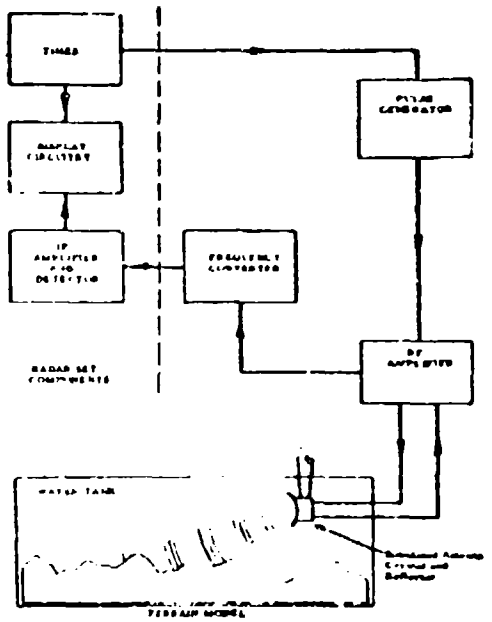


Figure 1

Aircraft motion is simulated by moving the crystal in X, Y and Z over the map surface. The AN/APQ-T2A, manufactured by both American Machine & Foundry and Otis Elevator, is a typical example of a simulator based on this technique.

Optical analog techniques in which light is radiated rather than ultrasonic energy have found application in several simulator programs. There are two basically different techniques which may be called optical analog. In the first of these, the terrain model is illuminated by a light from a flying spot scanner tube. A sweep on the face of the tube is imaged onto the map surface according to the radar range setting, and the de-

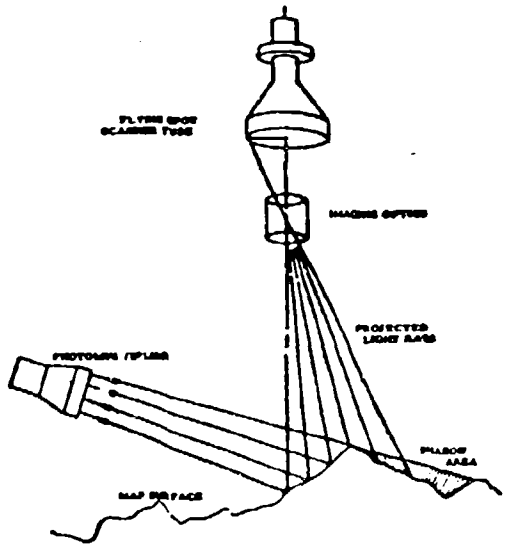


Figure 2

flection coils of the scanner tube are used to rotate the sweep in azimuth in accordance with the scanning mode of the radar antenna. The light reflected from the map surface is picked up by a photomultiplier tube located at the simulated radar antenna position. From the output of the photomultiplier tube, the resulting video is then suitably processed for display on the radar indicator. Figure 2 illustrates the details of this technique. It is presently employed in the Navy's 1525 radar trainer developed by Fairchild Astronics Division of Fairchild Engine & Airplane Company.

The second type of optical analog simulator reverses the positions of the illuminating source and detecting device. Here, a small high power-

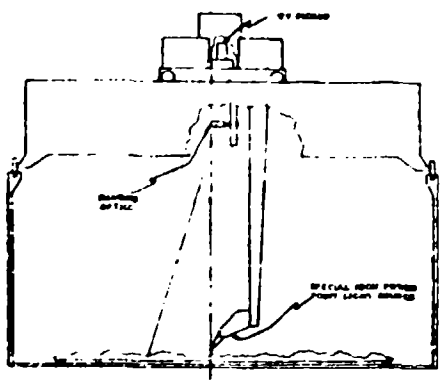


Figure 3

ed point light source is used to illuminate the map. The resulting highlights and shadows are then viewed from above, as shown in Figure 3, by a TV camera tube, such as an Orthicon or Vidicon which is scanned in PPI fashion. The output of these tubes then provides the video signals to the display. This technique is currently employed for generating land mass radar signals in the Air Force F10S-D flight simulator, currently in production by ACF Electronics.

It has also been employed, with the substitution of a photographic camera for the TV camera, for the preparation of reference films for use in "map-matching" missile guidance systems.

A third type of analog technique exists which may be best described as an electromechanical analog. Here, the terrain map is a specially prepared printed circuit contour map which is overlayed with a semi-conductive epoxy resin. When discrete voltages are applied to the copper contour lines in relation to the terrain elevation they represent, the semi-conductive epoxy produces interpolation voltages between the contour lines. Thus, the flat map surface depicts the terrain elevation in terms of different voltages. The "antenna" used in this technique consists of a series of contact fingers, arranged linearly along a member equal in scale size to the radar range. The voltages appearing across the contact fingers are sampled at the sweep rate and provide an accurate profile of the terrain for each successive sweep. Such a simulator has been developed by Cornell Aeronautical Laboratories of Buffalo, New York, under contract to the Air Force as a research tool for use in evaluating terrain avoidance systems.

To use such a technique for providing video signals from land masses, additional computer circuitry would be required. Also, to provide the simulation of returns from cultural targets, it would be necessary to encode this information on the map surface by some means.

The computational techniques for land mass simulation are based on providing a means for storage and readout of the terrain elevation information and the radar reflectivity information so that these data may be operated on electronically to compute the highlights and shadows which would occur, along with various other video signal effects. The first of these is the so-called "two transparency" technique exemplified by simulators presently in use or under development by the Navy, Air Force and RAF. One of the leading investigators of this technique has been the Curtiss-Wright Corp. Figure 4 illustrates the principles involved in this technique. Here the information is stored on two separate, carefully registered, photographic transparencies or glass plates, encoded in the form of density variation. At each point on the one plate the photographic density is a function of the terrain elevation and on the second transparency the density is a function of the radar reflectivity of natural and cultrual targets. When the scanning beam from a flying spot scanner is

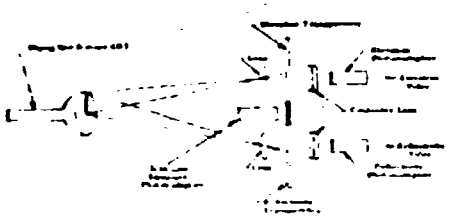


Figure 4

imaged onto the transparency at the appropriate scale, the transmitted light is intensity modulated in accordance with the coded information encountered by each sweep. This light is picked up by a photomultiplier tube whose output provides voltage analogs of the terrain elevation and radar reflectivity data. In order to derive meaning from this data, highlights and shadows must be computed along each sweep line in real time so that the appropriate video signals may be developed. Highlights and shadows are computed from the photomultiplier outputs representing terrain undulations and target reflectivities. To perform this computation it is also necessary to determine terrain slope (in order to get the highlights) and to enter into the computation the aircraft altitude, ground range and aspect angle. It is important to note here that the scale of these transparencies can be very much smaller than the smallest of the scales used for the terrain models employed in the analog systems. This results from the fact that the accuracy and resolution achievable by the photographic process are much greater than those of the terrain model making art.

Other variations of this basic technique use two reflective maps rather than transparencies, or a single transparent or reflective map which contains both reflectivity and terrain elevation information encoded in a special way so that the two types of data may be read out simultaneously.

These then are the basic simulation techniques in use today. In order to render an objective evaluation of each of them, we must consider the abilities of each to meet a wide variety of requirements.

Techniques Versus Requirements - What are the demands placed on a simulation technique?

Let us first have a look at one general problem common to all techniques. It is an important problem which becomes more severe in meeting the requirements for newer long range, high speed weapons systems. Navigational and presentational errors and distortions occur when the spheroidal earth's surface is represented on a flat plane. Of the many different map projections which may be used, the two systems of world mapping used by the military are the most practical for simulator application. These are the Lambert Conic Conformal projection used by the Air Force, and the Universal Transverse Mercator projection used by the Army. The Lambert Conformal system divides the earth's surface into 4° bands of latitude. A series of cones whose surfaces intersect the upper and lower parallels of each band are constructed for projection. When the projections of two adjacent bands are abutted they touch at only one point; the arcs describing the parallels of latitude are not concentric from one band to the next. Thus gaps exist when two or more bands are depicted in a single map.

In the Mercator system the earth's surface is projected in 6° longitudinal zones. When these zone projections are abutted they join only at the equator. The gaps increase in size with increasing north or south latitude. Thus, in representing the earth's surface on a flat plane, simulated missions over large areas of the earth's surface result in serious navigational errors.

Another related problem is the finite limit in the length of a mission which can be covered by a single map, no matter how small the scale. Problem interruptions for map changes reduce the effectiveness of the simulation and ought to be avoided.

There is one interesting solution to both of these problems which deserves mention here. The only way to accomplish uninterrupted global missions in a scale large enough to provide sufficient accuracy, in a manner which avoids the interruption in picturing data caused by gaps, and without suffering the distortions of planar projection, is through the use of a global relief map. A simulator, in which the entire earth is depicted in relief, has been the subject of an extensive design study by the Systems Division of Bendix Aviation Corporation and Aero Service Corporation. An artist's conception of the simulator is shown in Figure 5. In addition to simulating global flights of manned aircraft, a simulator based on the use of an accurate giant globe in relief could also simulate missile trajectories, satellite orbits and ground radar nets. Training missions could be set up to solve synthesized problems in radar navigation, terrain avoidance, radar bombing, search and tracking.

In addition to these principal applications the simulator, by virtue of its unique design, could synthesize problems in global dead reckoning navigation, photographic reconnaissance, television reconnaissance and pictorial telemetry, infrared reconnaissance, countermeasures and jamming, missile guidance-target homing, anti-missile and satellite reconnaissance.

The globe pictured in the illustration is an 84' diameter geodesic sphere fitted with a geoidal skin depicting the earth's surface in relief at a scale of 1"500,000 with a vertical exaggeration of 3:1.

The map scale has been chosen to provide a unit target area resolution of 250 feet (0.006" at 1:500,000 scale).

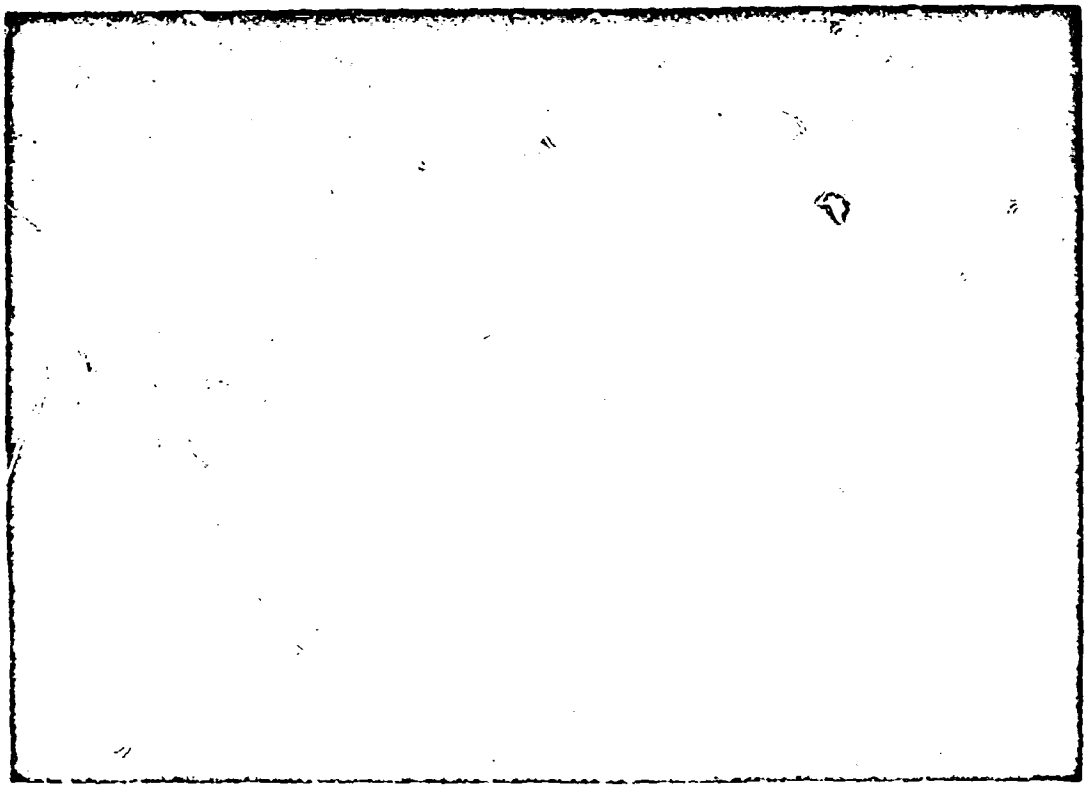


Figure 5

Let us now examine the criteria or yardsticks against which to evaluate the various simulation techniques. In judging a given technique, it is necessary to consider the achievable resolution, range of simulated altitudes, maximum ground range, overall equipment size, overall equipment complexity, maximum mission range, navigational and presentational distortions and errors, the capability of a terrain avoidance presentation, and ease of providing radar realism effects such as attenuation with range, azimuth beamwidth distortion, and cardinal point effects.

Features and Limitations of Existing Techniques

The plus and minus factors for each of these techniques may be conveniently tabulated and are shown in Figure 6. Some additional comment is helpful, however, since the reasons for and extent of some of these limitations should be noted.

On the plus side of the ultrasonic technique is its long history of continual development and improvement. The number of simulators presently deployed in the field attests to the fact that this system works. The scale ratio between the propagation of microwave energy in air and the propagation of ultrasonic energy in water restricts the scale of the terrain model used to 1:200,000. There are both advantages and disadvantages to this large scale. On the plus side, it means that the navigational and presentational distortions and errors due to the map projection are small. This larger scale also means that the overall accuracy of simulation is high because of the accuracy with which a map at this scale can be constructed. On the other hand, this large fixed scale means that the maximum mission range is extremely limited with respect to modern mission requirements unless a large map is used. Thus, the overall equipment size is quite large. For simulating B-47 flights in an ultrasonic trainer, for example, it was necessary to fabricate a terrain model measuring 9' × 36' in a single piece.

While the overall equipment size may be large, it is not exceedingly complex. The electronics and the mechanical design are relatively straightforward since no computational circuitry is necessary. Highlights and shadows and most other video signal effects occur naturally in this system because it mimics the principles and geometry of the actual radar. Because of this limitation in principle, it is conceivable that a terrain avoidance capability could be added to an ultrasonic simulator in the same manner that

Technique	Parameters which limit resolution	Minimum Altitude	Maximum Ground Range	Relative Overall Size	Relative Complexity	Maximum Mission	Navigational & Presentational Errors & Distortions	Terrain Avoidance	Video Signal Effects		
									Azimuth Beamwidth	Attenuation with range	Cardinal point
Ultrasonic	1. Map (res. @ 1:200,000 is 100') 2. Large crystal size	5000' (AN/APQ-T2A)	100 mi.	Large	Simple	32.88 n. mi. / foot of map @ scale 1:200,000	Relatively small due to large map scale	Not on present models due to minimum altitude limitations.	Yes	Yes	Yes-with special map surface treatment.
Optical-flying spot scanner & photomultiplier pickup	Map (res. @ 1:500,000 is 250')	2500' (1525)	Limited by light output of flying spot scanner. Presently about 120 miles with 1:500,000 scale map.	Large	Simple	82.2 n. mi. / foot of map @ scale 1:500,000	Somewhat more than ultrasonic due to smaller map scale	Not on present models. Feasibility under study.	Must be synthesized	Yes	Yes-with special map surface treatment.
Optical-point light source and TV pickup	Map (res. @ 1:500,000 is 250');	100'	Limited by light output of point light source. Presently 150 miles with 1:500,000 scale map.	Large	Simple	82.2 n. mi. / foot of map @ scale 1:500,000	High targets subject to relief distortion error. Same map distortion problems as above unless spheroidal map is used.	Yes-with auxiliary storage & readout or with scanning pencil beam light source	Must be synthesized	Yes	Yes-with special map surface treatment.
Computational	Photo plate resolution (res. = 150' @ 1:1,000,000 scale & 60 lines/mm res.) FSS spot size may limit on long range storage.	No limit	Not limited by the technique. Can accommodate present and future airborne radars.	Small	Involves computer circuitry	822 n. mi. / foot of map @ scale 1:5,000,000	Relatively high due to small map scale	Yes	Must be synthesized	Yes-computed function	No-feasibility under study

Figure 6

the actual radar performs this operation. Unfortunately, however, such a modification would be pointless because the physical size of the radiating crystal is so large as to limit the minimum altitude (which can be simulated) to a few thousand feet above the terrain. This is perhaps the most severe limitation of this technique in the light of the present tactical requirements for low level flight. It does not mean, however, that the technique is not usable in the simulation of radars and systems where low altitude is not a problem. Another limitation is the 100-mile maximum ground range due to the attenuation of ultrasonic energy in water. Such a restriction rules out the use of this technique in simulating long range radars.

Additional maintenance on ultrasonic simulators is necessary because of the large water tank. Problems attendant on maintaining proper water temperature, purity, lack of air bubbles, tank cleanliness, etc., are a nuisance but are not technical limitations of the technique.

In both the optical analog techniques as well as in the ultrasonic technique the resolution, accuracy and maximum mission range are determined and limited by the map scales used. Recent work in relief map techniques make it possible to achieve 500' horizontal resolution at a scale of 1:1,000,000. A standard maximum vertical error is 0.030" and it is possible to achieve 0.020". As the map scale increases, resolution and accuracy increase, but the maximum radar range is limited by the light output of the flying spot scanner in one case and the point light source in the other case.

One of the principal differences between the two techniques is in the low altitude limitation. In the flying spot scanner-photomultiplier technique, the physical size of the multiplier tube restricts the minimum altitude to about 2,500 feet. However, recent developments in high power point light sources allow the use of the point light source-TV pickup technique for simulation at very low altitudes. A typical instrument designed for use as the optical antenna in an analog simulator is shown in Figure 7. Modification programs or product improvement programs are currently under way to improve the low altitude performance of the optical technique using the flying spot scanner and photomultiplier pickup.

Another difference between the two optical analog techniques is that relief distortion occurs in the point light source-TV pickup method, but not in the flying spot scanner-photomultiplier pickup method. The placement of the TV camera violates the radar geometry in order that the shadows can be seen. Viewing the illuminated scene from above, the camera sees the terrain peaks in perspective rather than orthogonally and thus depicts them with a radial displacement outward from the optical axis. The extent of displacement is a function of the ground range, the height of the peak and the height of the camera. Figure 8 shows the relief distortion expressed in percentage of ground range for objects of in-

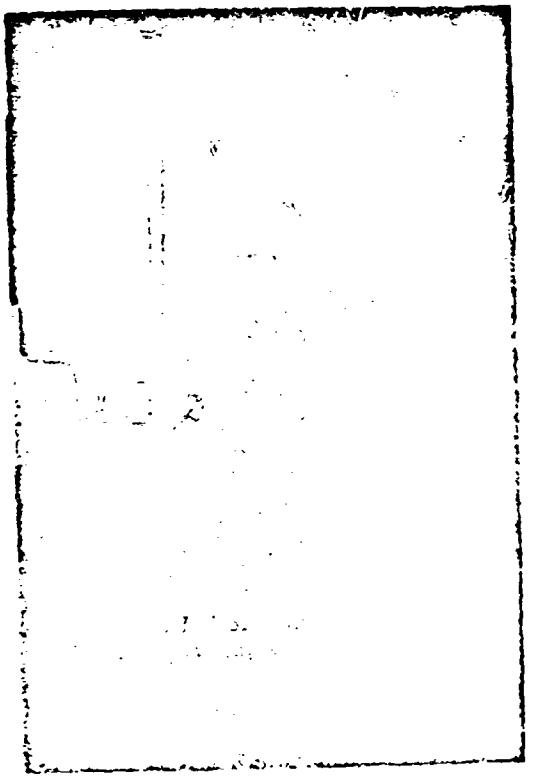


Figure 7

creasing height at several vertical exaggerations of a given horizontal map scale. The camera height is constant in these calculations. Since the maximum camera height is usually fixed by mechanical design considerations this error can best be minimized by using small scale maps having minimal vertical exaggeration. The actual ratio of the vertical scale to the horizontal depends on the particular geographic area covered by the map.

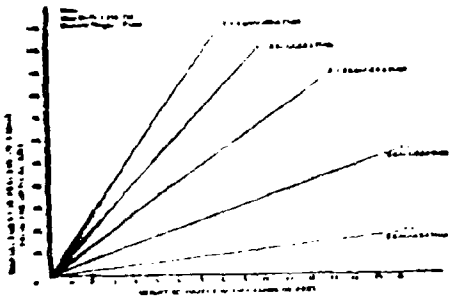


Figure 8

A plus factor on the side of the optical techniques is that the overall equipment, though large in size, is relatively simple in mechanical design and electronic circuitry. Existing simulators using these techniques have met their requirements, for the most part, with excellent results.

The principal assets of the computational techniques are the tremendous reduction in map scale allowing extended missions over several thousand miles, and the absence of a low altitude limit. The variation in radar return and video signal effects with altitude is a computed function. Likewise, shadows, terrain highlights and returns from shore lines and other natural and cultural features (which may be dependent upon horizontal and vertical aspect angles) must be computed. This makes the resulting electronic circuitry of the overall equipment very complex and gives rise to the requirement for maintenance personnel skilled in computer circuitry. On the other hand, personnel skilled in mechanical maintenance are required to maintain optical analog equipments.

It should be noted, however, that the use of map scales of 1:1, 000, 000 or 1:5, 000, 000 creates great problems in preserving planimetric accuracy and resolution of the system. It is obviously difficult to store highly detailed information with fidelity at these scales. An even more serious problem is that of maintaining the microdensity of the photographic image at the proper level to represent the terrain elevation. Unfortunately, there is not much published information on the number of reliably reproducible altitude increments that can be represented and read out of a photographic transparency at these small scales. The elevation accuracy for purposes of determining shadow lengths and for terrain avoidance presentation is limited by the ability to encode the terrain elevation information properly, to read it out and to compute it. Extensive programs are now under way to improve the photographic techniques involved in preparing maps of this small scale in order to minimize these problems.

Terrain avoidance capability, a recent requirement, is managed rather simply in the computational technique of simulation. As noted on the chart, it is possible to provide a terrain avoidance capability for the optical analog technique using the point light source and TV pickup. Such a modification would be feasible but difficult at small map scales since high precision mechanical components are required.

A word should be said regarding hybrid systems which attempt to combine the best features of several techniques. An example of such a system is the F105 Simulator in which the resolution and accuracy of the optical technique are accompanied by a system using a reflective map on which is encoded terrain elevation information. The purpose of the auxiliary map, the readout for which is synchronized with the position and motions of the main unit, is to provide information for a terrain avoidance presentation.

Thus, it is seen that no one technique provides all the answers. Each has its disadvantages and these are not all similar.

Present Directions and the Future

Current tactical weapons system requirements are for a low altitude capability and a terrain avoidance capability. As noted in the earlier discussion, the newer computational techniques lend themselves directly to meeting these two requirements. The successful employment of these techniques, however, is not without some sacrifice in accuracy and resolution. High performance aircraft flying at higher speeds for longer ranges definitely require a longer mission capability than previously designed into the analog type simulators. A look at the future, however, indicates that newer radars going into these high performance weapons systems will require in their simulation, greater accuracy and resolution in both the bomb-nav presentation as well as in the terrain avoidance mode. The importance and stringency of each of these requirements will vary, however, for each individual weapon system. Requirements for Air Force bombers will be different from those for Navy heavy attack bombers; fighter aircraft will have yet another set of requirements.

It would be wasteful of both time and money to expect each of the using agencies to invest in the development of a simulator which did anything more than meet their specific requirements. A universal technique does not now exist and it seems quite clear that if it were possible to develop one the value of such versatility would be outweighted by the extreme cost. No technique cited in this paper is without its limits. Each has something to offer.

It can be concluded then that a simulation system should be selected on the basis of a design which meets the particular requirements rather than on the basis of a technique which is new and promises "pie in the sky". Programs directed toward advancing the state of the art should be fostered, but should be carried on separately from those which are meant to provide workable simulators for a going weapon system. It is therefore proposed that restraint and objectivity be exercised in selecting a technique to meet a particular need.

Bibliography

1. Levine, Daniel, Radargrammetry, New York, McGraw-Hill, 1960.
2. Cornell Aeronautical Laboratory, Report No. CM-776-T-66, A Terrain Avoidance Display Evaluation Simulator.
3. Slattery, J. T., and Kamentsky, M., "Synthetic Representation of Terrain Features on a Simulated Airborne Radar Display", IRE Transactions on Military Electronics, July, 1959, MIL-3, 75-82.
4. "Utilization Handbook for Bombing AN/ASB-1A Advanced Operator Training Device 15-Z5, "NAVEXOS P-1626, U. S. Naval Training Device Center, Port Washington, N. Y., 1958.
5. "Handbook Service Instructions Radar Training Set AN/APQ-T2A", T. O. 43D7-4-3-2, Air Force, DPS, Ogden, Utah, 1959.
6. "A Global Simulator", Aero Service Corporation, Philadelphia, 1958.
7. "Final Report on High Power Point Light Sources", Report File 181, 15 January 1959, Westgate Laboratory, Inc., Yellow Springs, Ohio.

OPTIMUM CAPACITOR CHARGING EFFICIENCY FOR SPACE SYSTEMS*

Philip M. Mostov, Joseph L. Neuringer and Donald S. Rigney

Plasma Propulsion Labs., Republic Aviation Corp., Farmingdale, New York

SUMMARY

In space systems the efficient utilization of energy can be critical. Many pulsed operation devices, e.g., a plasma engine, involve charging a condenser periodically. The efficiency of energy transfer (η) to an initially uncharged condenser (C) when a constant DC source voltage is applied through a resistance (R), with inductance (L) assumed zero, is shown limited to 50%, even if R varies arbitrarily during the charging process. If $L>0$, η can be made to approach 100% by charging in a periodic mode and terminating at the end of the first $\frac{1}{4}$-cycle. The requisite L, which is a function of R, C and charging time T, can be formidable and its weight large. To provide guide lines in the selection of practical voltage shapes, the Calculus of Variations is used to derive theorems for the "perfect" time-shaped applied voltages that optimize η when the delivered energy, L, C and T are fixed. Four modes of prescribing R, as a function of time (t) and/or current (i), are treated: a. Constant R; the key condition is constant current for the full charging time allowed; the voltage is a modified "elevated ramp"; $\eta_{opt}=1/(1+2RC/T)$, can approach 100% and is independent of L. b. R(t). c. R(i). d. R(t,i). Initial condenser voltages can lead to improved efficiencies, and are compatible for use with a plasma accelerator stage where they tend to imply large accelerator energy utilization factors. Promising voltage shapes, approximating the ideal, are synthesized.

I. INTRODUCTION

In an ever increasing number of power systems, as well as those combining power and communication functions, it is of paramount importance to optimize the efficiency of energy transfer η, i.e., the ratio of the energy delivered to that given up by the source. Such an optimization procedure is distinctly different than the usual one in most communication systems, where a maximum absolute power transfer is generally aimed for without much regard for the efficiency. In fact, in the usual maximum power transfer situation, η is only about 50%.

Unfortunately, a loss of energy is generally reflected in the weight of the energy source. For naval or even ground applications this weight can often be somewhat of a problem, but for aeronautical and especially astronautical applications it can be quite crucial in its effect on the total system. This is especially true for those propulsion systems in which most of the energy is supplied to the propellant rather than being generated in the propellant itself[1].

Many pulsed operation devices, e.g., Republic's plasma engine[4,5], involve charging a condenser periodically. A simplified schematic of a typical LCR series feeding network is given in Fig. 1. The energy source S charges up the capacitor bank C relatively slowly by means of the current i(t) through the resistance R(t) and the inductance L, where t denotes the time. In such a stage the capacitor bank plays the role of a load. The bank voltage and charge are v(t) and q(t) respectively, and the source voltage e(t). At the completion of the charging time, T, the bank discharges in a relatively short period of time, i.e., it discharges in a pulse fashion, into the next stage.

Kirchoff's law applied to Fig. 1 leads to

$$e(t) - i(t)\,R(t) - L\frac{di}{dt} - v(t) = 0; \qquad (1)$$

where

$$v(t) = v_0 + \int_0^t i\,dt, \qquad (2)$$

v_0 being its initial value. If a particular e(t) is applied to a given LCR with specified initial conditions, all the subsequent behavior is determined, e.g., i, v, q, η as well as the energies supplied by the source, consumed irreversibly by the resistance, stored in the magnetic field and delivered to the condenser.

Both the initial current and v_0 will be taken as zero, except the latter will be allowed to be different than zero in part IV.

---

*Abridged from Refs. 1-3, which were partially supported by the Office of Naval Research, and should be referred to for more detail. Numbers in parentheses refer to the reference section.

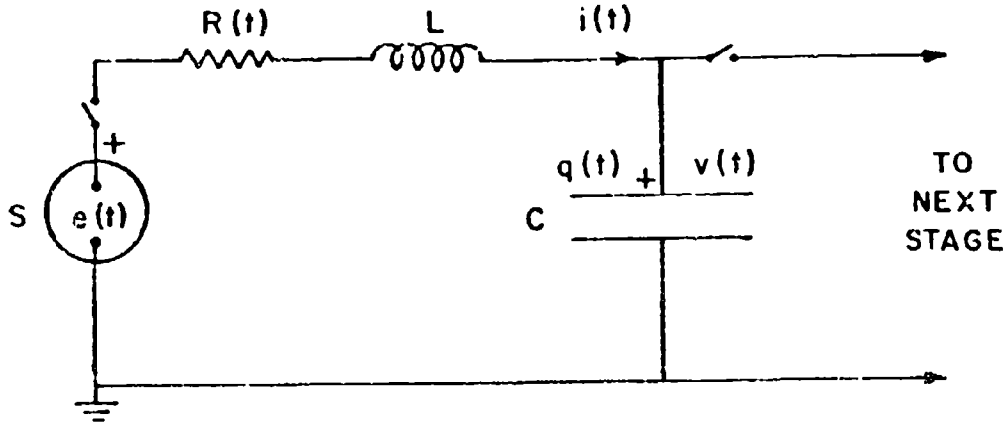


FIG. 1. SCHEMATIC OF THE CHARGING CIRCUIT

II. CONSTANT DC APPLIED VOLTAGE

The case where $e(t)$=constant, a common well-known arrangement, can serve as a useful reference case, as well as help illustrate the basic ideas.

If R and C are constant, and if additionally L=0, η is well known to be

$$\eta_\infty = \tfrac{1}{2}(1 - e^{-T/RC}),$$

a monotonically increasing function of the single parameter T/RC, i.e., the number of RC time constants used during the charging period. η vs. T/RC for this case is plotted in Fig. 2, in which several other cases to be discussed below are also included for comparison. The maximum value of η attained for the present case is 50%. That such a 50% limitation can make the capacitor charging stage a serious weak link in many systems has been emphasized by several writers[6].

This 50% limitation can be shown to govern even if R varies arbitrarily[1] during the charging process. Thus constant voltage schemes that attempt to circumvent this inherent 50% limitation by varying R are doomed to failure, e.g., those constant current generators that use a constant voltage power supply and vary the effective internal resistance. However, as shown below, if the constant current is attained by tailoring the source voltage this 50% limitation no longer necessarily applies - in fact almost 100% efficiency is possible in some cases.

When $L > 0$[2], η can be made substantially larger than 50% under some conditions. As is well known, the current response mode can be aperiodic, critical or periodic. For the first two, it can be shown that $\eta \leq 50\%$. For the periodic mode, however, η can exceed 50% under appropriate conditions. In particular, the maximum η is attained by arranging to cut off at the end of the first $\frac{1}{2}$-cycle. This η_{max} is given by

$$\eta_{max} = \tfrac{1}{2}\left(1 + e^{-\frac{1}{2}T/(L/R)}\right), \qquad (3)$$

which can be expressed in terms of the single parameter T/RC since

$$\frac{T}{L/R} = \frac{2\pi^2}{\frac{T}{RC} + \left\{\left(\frac{T}{RC}\right)^2 - \pi^2\right\}^{\frac{1}{2}}} \qquad (4)$$

must hold ($\frac{1}{2}$-cycle time condition). The L required is determined for a given R, C and T; it often can be quite large and its associated weight formidable in some cases, especially if its own resistive contribution to the total resistance is to be kept reasonable.

There are two branches of $\eta(T/RC)$ (cf. Fig 2, on which Eq. (4) is also plotted to help indicate the required associated L's). On the higher branch of $\eta(T/RC)$, to which the lower $T/(L/R)$ branch and therefore the higher L correspond, η increases from 52.2 to 100%; on the lower branch, η decreases from 52.2 to 50%. Note that Eq. (4) can only be satisfied for $T/RC > \pi$ if $T/(L/R)$ is to be real. If Eq. (4) is not satisfied, Eq. (3) does not govern, but substantial η's are still possible provided $T/RC > 4\pi/(L/R)$ (oscillatory condition).

Though fixed source voltage schemes can lead to high η's in some cases and ranges, it may not be practical to use such a scheme, there often being inherent losses in both its creation and application, as well as critical timing requirements.

III. OPTIMIZATION THEOREMS

In order to serve as a guide in the selection of voltage shapes that are both efficient and practical, a series of theorems on the optimization of the efficiency have been derived. Employing the techniques of the Calculus of Variations[7,8], the "perfect" time-shaped applied voltages that optimize η for prescribed LRC, T and energy delivered have been found. Both a constant R, as well as a variable (function of t and/or i) one, have been treated. The details of the application of the Calculus will not be given here. It turns out to be simplest to recast the original question on the applied voltage to one on the associated current. In either case, the answer is most succinctly stated in terms of the current-time characteristic.

A. R CONSTANT

For constant R, the key condition is found to be a constant current lasting the full time T allowed for the charging process. In terms of the final charge on the condenser, Q, the energy on the capacitor is simply $Q^2/2C$. Since the energy and capacitance are prescribed, it follows that the needed Q is determinable in advance (except for sign). Since $q = \int i\,dt$, constant current means that the current is given by

$$i = \frac{Q}{T}$$

(Fig. 3). The i^2R loss rate is also constant: $i^2R = Q^2R/T^2$. A constant current feeding the condenser means that both q and v rise linearly in time from 0 to respectively Q and V, V denoting the final condenser voltage: $v = Vt/T$. Δe, the iR drop, is also constant: $\Delta e = RQ/T$.

Since $e = v + \Delta e + L\,di/dt$, when $L=0$ the required applied voltage is always Δe higher than v, and of course does not pass through the e-t origin as one might suppose:

$$e(t) = \frac{V}{T}(RC + t).$$

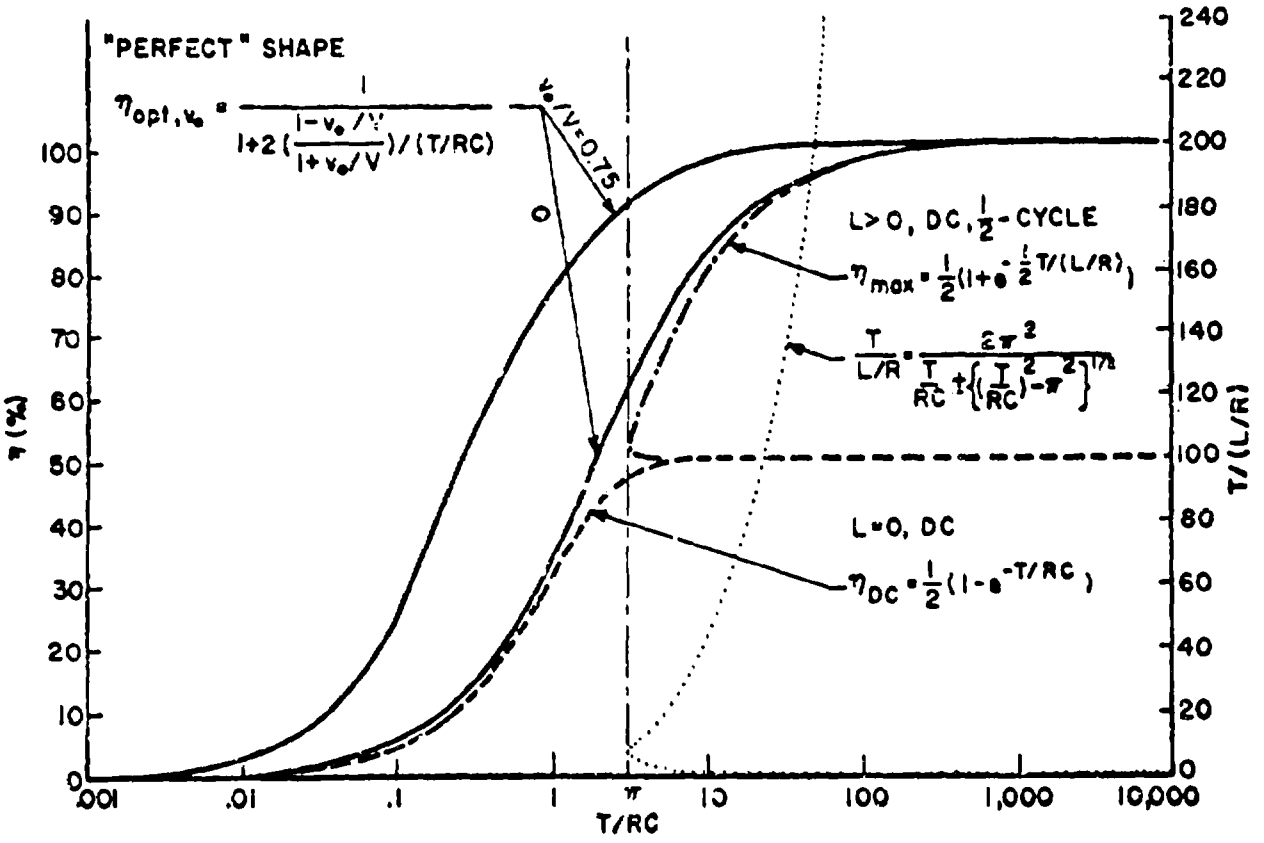


FIG. 2. η vs. T/RC FOR

$L=0$, DC
$L>0$, DC, $\frac{1}{2}$-CYCLE
ANY L, PERFECT SHAPE; $v_o/V=0$, 0.75.

ALSO T/(L/R) vs. T/RC AT THE $\frac{1}{2}$-CYCLE TIME FOR THE $L>0$, DC, PERIODIC MODE.

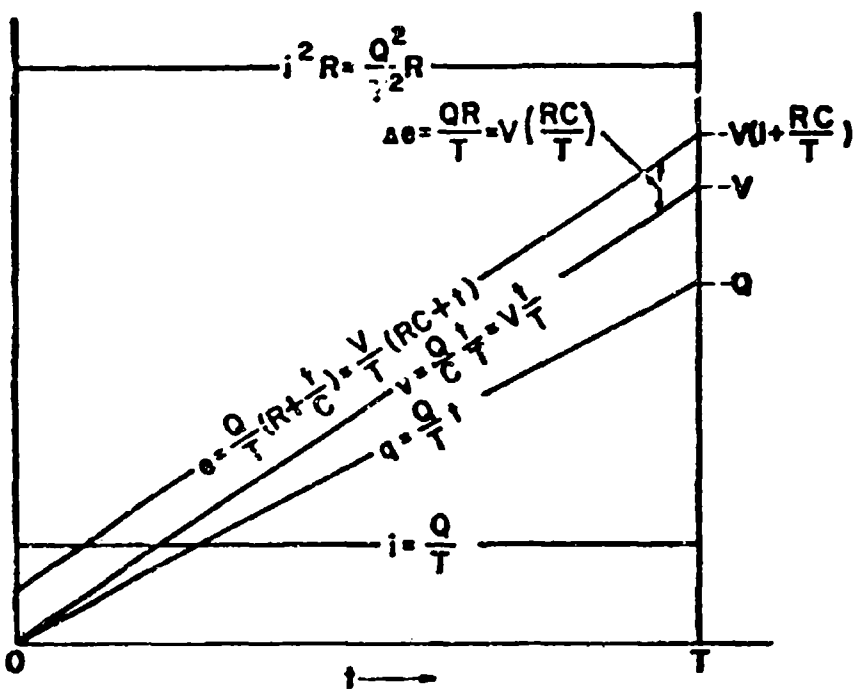


FIG. 3. i, q, v, Δe, e AND i^2R vs. t FOR OPTIMUM γ, WHEN THE DELIVERED ENERGY, R, C AND T ARE PRESCRIBED CONSTANTS AND L=0. WHEN L>0, SUITABLE POSITIVE AND NEGATIVE PULSES MUST BE SUPERIMPOSED ON e AT t=0,T RESPECTIVELY.

When $L > 0$[2], suitable positive and negative source voltage pulses of infinite amplitude and infinitesimal duration are superimposed at the beginning and end, respectively, of the charging time, in order to bring the current up from zero to Q/T and down to zero again.

The optimum η achievable is found to be

$$\eta_{opt} = \frac{1}{1 + \frac{2}{T/RC}}$$

(Fig. 2) which increases monotonically with T/RC and approaches 100% ! This η is independent of the value of L that a particular configuration might have, i.e., no special L is required.

Thus we have shown that we can overcome some of the restrictions, and the associated weight problem, on the inductance by suitably tailoring the applied voltage shape. Neither the minimum irreducible L (which, when a DC is applied usually means an aperiodic response and therefore a 50% efficiency limitation), nor the large L that may be needed for compatibility with other facets of a given design, need be a direct limitation on the η attainable, i.e., η_{opt} depends only on T/RC.

Though η_{opt} exceeds the upper branch of η_{max} by only a small amount in some ranges of T/RC, especially at the higher T/RC's necessary for substantial efficiencies, it should be kept in mind that the requirement of a large L for the $\frac{1}{2}$-cycle condition can become quite severe and untenable for many space designs.

B. R VARIABLE[3]

In many cases a variable resistance is more realistic than a constant one since the effective R can vary due to a number of causes, e.g., heating, cooling, effective radiation resistance, effective internal resistances (including those of constant current generators), skin effects, solid state effects, plasmas, discharges, non-linear components and circuits.

If the resistance is a prescribed function of time, $R(t)$, the key condition for optimum efficiency is found to be the constancy of the iR product:

$$i(t)R(t) = \frac{Q}{T} R_H(T)$$

a constant; where $R_H(t)$ is the harmonic time average of $R(t)$ over the interval 0-t defined by

$$\frac{1}{R_H(t)} = \frac{1}{t}\int_0^t \frac{1}{R(t)}\,dt,$$

i.e., a time average reckoned by reciprocals.

The optimum η is found to be

$$\eta_{opt} = \frac{1}{1 + \frac{2}{T/R_H(T)C}},$$

the same form as for the constant resistance case but with R replaced by $R_H(T)$, the harmonic average over the interval 0-T. Consequently the discussion of η_{opt} is somewhat similar to the constant R case, and large efficiencies are generally possible. An important difference, however, is that $R_H(T)$ is a function of T for prescribed $R(t)$, so that T and $R_H(T)$ are not independently specifiable in attempting to achieve favorable values of $T/R_H(T)C$.

$v(t)$ and $e(t)$ are also more complicated:

$$v(t) = V\frac{R_H(T)}{R_H(t)}\frac{t}{T};$$

$$e(t) = \frac{VR_H(T)C}{T}\left(\frac{1}{R_H(t)C}t + 1 - \frac{L}{R^2(t)}\frac{dR(t)}{dt}\right),$$

with additional suitable positive and negative source pulses at $t=0,T$ if $L > 0$.

The resistance a prescribed function of current, $R(i)$, is usually a closer approximation to the actual physical phenomenon than $R(t)$, i.e., the value assumed by a resistance usually depends more on the current through it rather than on the particular time that it happens to be, though of course it is actually a function of both. For a wide class of $R(i)$'s the key condition and η are found to be similar to that of the constant R case. The resistance a prescribed function of both time and current simultaneously, $R(t,i)$, has been also treated.

In future papers, R a prescribed function of its previous current-time history, as well as several allied modes of specifying R, L and C, will be treated, e.g., the commonly encountered phenomenon of L varying with i. The effect of non-zero initial and final currents will also be given.

IV. INITIAL CONDENSER VOLTAGE

The effect of an initial voltage on the condenser has been considered for both DC and perfect-shaped applied voltages. Improved efficiencies, approaching 100% when the initial voltage approaches the final voltage,* are possible. However, under many conditions of normalization, the capacity, and its weight, also increases.

For a constant prescribed R, the optimum efficiency turns out to be

$$\eta_{opt,v_0} = \frac{1}{1 + 2\left(\frac{1 - v_0/V}{1 + v_0/V}\right)/(T/RC)} \quad . \quad (5)$$

The corresponding needed applied voltage is

$$e = v_0 + \frac{V - v_0}{T}(RC + t)$$

(an elevated ramp) plus suitable pulses superimposed at $t=0$ and T if $L>0$.

As the parameter v_0/V increases from -1 to $+1$, η_{opt,v_0} increases monotonically from 0 to 100%; so that depending on whether v_0/V is greater or less than zero, η_{opt,v_0} is respectively greater or less than the η_{opt} corresponding to $v_0/V=0$. η_{opt,v_0} is plotted vs. T/RC in Fig. 2 for $v_0/V=0$ and 0.75. For $T/RC=3$, it jumps from 60% when $v_0/V=0$ to 90% when $v_0/V=0.75$, an improvement of a factor of $\frac{1}{2}$ which can be very helpful in many tight, weight-conscious space designs. Should it be desirable, or necessary, to operate at lower values of T/RC, the fractional increase of η_{opt,v_0} is even more dramatic, e.g., at $T/RC=0.1$, the fractional increase exceeds 5. Eq. (5) illustrates that if $v_0=0$, operating at $T/RC<3$ implies that even the optimum efficiency of energy transfer will be limited to values below 60% which in many cases is impractical except for relatively short missions when the weight of the power supply tends to be less important than that of the capacitor bank. For the longer missions, however, when the weight of the power supply often tends to be more important than that of the capacitor bank, one can still operate with high efficiency at $T/RC<3$ if a substantial initial voltage is used.

*In the type of plasma engine being developed at Republic, the initial and final voltages of the accelerating stage are identical to the final and initial voltages, respectively, of the charging stage - the condenser is a common one. Fortunately, a high ratio of initial to final voltage in the charging stage not only increases the charging efficiency, but, as shown in Ref. 9, it also tends under many operating conditions to increase the energy utilization factor in the accelerating stage.

For any value of v_0/V, $\eta_{opt,v_0} \to 100\%$ as $T/RC \to \infty$. The formula shows that for $T/RC > 50$, the increase of η_{opt,v_0} due to the use of a positive initial voltage is not too significant since it already exceeds 90% when no initial voltage is involved, though for very long missions it can still be helpful.

Under some conditions of normalization,(1) e.g., when the energy actually used (as opposed to the total energy the capacitor bank is called upon to hold) is fixed and V and the energy densities in the capacitor bank are restricted to stay within given bounds, the C needed is a function of v_0. Under such circumstances, v_0 and C (and therefore also T/RC) in Eq. (5) would not be independently specifiable and a useful interpretation of the equation would have to be correlated with other relationships.

V. PROMISING PRACTICAL SHAPES

Employing the theorems given here as a guide, one can proceed to synthesize promising, and perhaps more practical, voltage shapes approximating the perfect one. Several typical such shapes (Fig. 4) are given below. Each resembles the ideal shape in one or more important features that probably have reasonable chances of being achievable, while minimizing those aspects of the ideal shape for which too much of a penalty may be incurred.

1. Ramp, i.e., a slope through the e-t origin.

2. Elevated Ramp, i.e., a slope not through the origin. Though it resembles the perfect shape for the constant R, L=0 case, it is not necessarily exactly identical to it since the initial value and slope may need to be adjusted to compensate for the fact that the requisite pulses are not used. If the delivered energy is prescribed, the values of the parameters specifying the applied voltage may be arranged so as to maximize the efficiency.

3. Elevated Ramp with Suitable Pulses of Finite Width and Height Superimposed, rather than the infinitely high and infinitesimally narrow pulses needed for the perfect shape in the constant R, L>0 case.

4. "Plateaued" Elevated Ramp, i.e., a voltage that starts out as an elevated ramp but eventually becomes flat. If the flat portion were continued long enough, the current would decrease to zero. Thus such a modified elevated ramp has the advantage of tending to minimize the cut-off current and the attendant wasted magnetic energy, without resorting to sharp pulses (perfect shape), as well as not being critically sensitive to the exact time of cut-off (½-cycle charging). The concomitant dividends of easing the

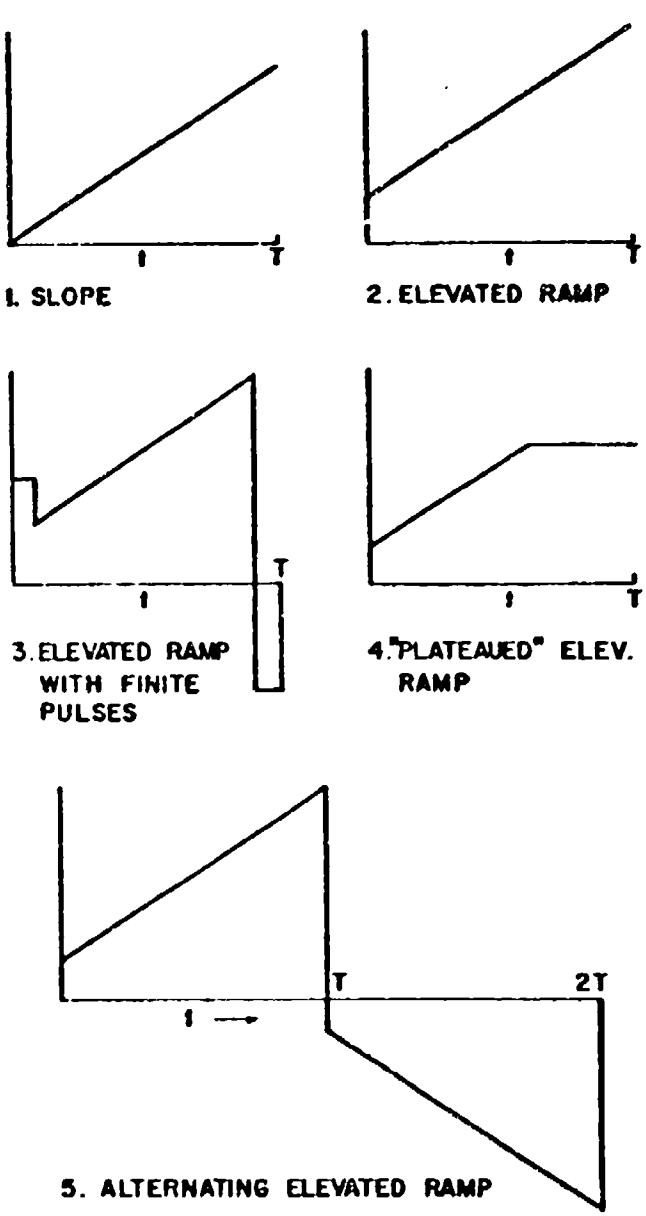


FIG.4 SEVERAL PROMISING VOLTAGE SHAPES

switch wear, tear and reliability problem, if one is used, can be important in some ranges of operation.

5. Alternating Elevated Ramp, i.e., for the first of each of two successive T periods the voltage is one polarity and for the second it is the opposite. (The short discharge time between these two charging periods is not shown). Note that for the second period, even though the polarity of v_0 is reversed, so is that of V, and the v_o/V ratio remains positive. Consequently, the same increased efficiency (compared to the $v_o/V=0$ case) is possible for both periods. Thus it is not necessary to restrict the source voltage to single polarity types, nor to require that the final and initial voltages of the pulse discharge into the load (e.g., a plasma accelerator) have the same polarity as each other, i.e., the voltage polarity at the end of a particular discharge does not have to be the same as that with which the discharge began. All that is needed is that the polarities of the initial and final voltages of a particular charging period be the same.*

The probably high efficiencies attainable with these and other promising shapes are being investigated both analytically and experimentally. If these shapes can be generated efficiently enough by modified conventional power supplies, a useful efficient generator-feeding network complex would appear to be easily achievable. Otherwise one of the benefits of the present study is to indicate a gap that novel, unconventional generators should attempt to bridge. Furthermore, an overall optimization of the combined generator, feeding network and plasma accelerator would appear to be a fruitful next step.

ACKNOWLEDGMENTS

The authors are pleased to thank A. E. Kunen, head of PPL, for inspiring this problem and his guidance. Thanks are also due Profs. L. Brillouin, H. Grad; Drs. L. Aronowitz, L. Kraus, H. Malamud; and J. Port; for helpful discussions and P. Chouteau for diligent typing.

*Fortunately, these requirements and possibilities are also helpful in a plasma accelerator stage, where it is desirable to have the absolute value of the ratio of the final to initial discharge voltages high (the energy utilization factor is a monotonically increasing function of the square of this ratio). Thus it turns out that the charging and plasma accelerator stages are highly compatible when a combined design is sought.

REFERENCES

1. Philip M. Mostov and Joseph L. Neuringer, "RC Charging Scheme to Optimize the Efficiency of Energy Transfer to a Condenser", May 9, 1959, PPL No. 104, Plasma Propulsion Labs., Republic Aviation Corp.

2. Philip M. Mostov and Joseph L. Neuringer, "LRC Charging Scheme to Optimize the Efficiency of Energy Transfer to a Condenser", June 8, 1959, PPL No. 106, Plasma Propulsion Labs., Republic Aviation Corp.

3. Philip M. Mostov and Joseph L. Neuringer, "LRC Schemes to Optimize the Efficiency of Energy Transfer to a Condenser for Prescribed Variable R", Sept. 7, 1959, PPL No. 113, Plasma Propulsion Labs., Republic Aviation Corp.

4. Alfred E. Kunen, "The Electromagnetic Pinch Effect Applied to a Space Propulsion System", August 4, 1958, Republic Aviation Corp.

5. Alfred E. Kunen and William McIlroy, "The Electromagnetic Pinch Effect for Space Propulsion", August 1958, PPL No. 115, Plasma Propulsion Labs., Republic Aviation Corp.

6. Bernard Gorowitz and Boyd W. Harmed, "Measurements of Velocity and Momentum with a Pulsed T-Tube Plasma Generator", Feb. 20, 1959, Technical Information Series, No. R59SD314, Missile and Space Vehicle Dept., General Electric, especially page 24.

7. Robert Weinstock, Calculus of Variations with Applications to Physics and Engineering, McGraw-Hill, New York, first edition, 1952.

8. Richard Courant and D. Hilbert, Methods of Mathematical Physics, Interscience, New York, 1953.

9. Philip M. Mostov, Joseph L. Neuringer and Donald S. Rigney, "The Theory of a Plasma Acceleration Process Suitable for Space Propulsion and Fusion Injection", Feb. 15, 1960, PPL-TR-60-5, Plasma Propulsion Labs., Republic Aviation Corp.

MATTS (Multiple Airborne Target Trajectory System)

By: Walter J. Zobbe, Cubic Corporation

Introduction

Effective evaluation of the air-to-air weapon systems requires accurate information about the interceptor aircraft, target drone, and missile trajectories, including the point of closest approach of the missile to the drone. For missiles with high-yield warheads, it is also necessary to provide interceptor data following missile launch in order to evaluate the escape maneuver. Several systems have been developed which provide simple scalar miss distance, but the system described in this paper is the first which can provide the trajectory and vector miss-distance information essential for complete weapon-system evaluation.

This new system, the AN/GSQ-29 (XI-1), is officially designated as "Scoring System, Vector Miss Distance," and is also known as MATTS (Multiple Airborne Target Trajectory System). It is installed on the Eglin Gulf Test Range in Florida. MATTS is an electronic tracking system whose primary mission is evaluating the proficiency of Air Defense Command pilots in the use of air-to-air rockets and missiles. Similar systems are presently or soon to be installed at the Naval Ordnance Test Station (MIDAS - Missile Intercept Data Acquisition System), White Sands Missile Range (ITS - Integrated Trajectory System), Eglin Gulf Test Range (DOETS - Dual Object Electronic Tracking System), and the Pacific Missile Range (Bi-COTAR Range Safety System).

MATTS is a ground-based c-w phase-comparison system. Each interceptor, missile and drone, carries a small c-w transmitter. The transmitted signals are received by ground stations at Cape San Blas and Carrabelle (38 miles apart) as indicated in Figure 1.

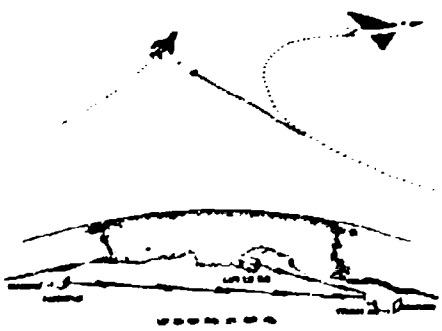
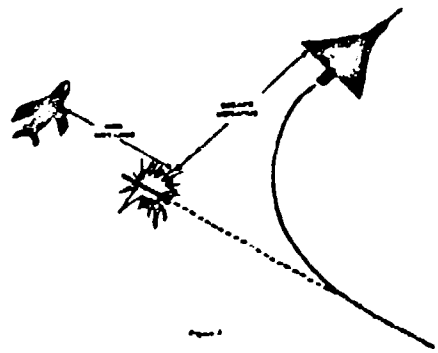


Figure 1

Each ground station determines a pair of direction cosines associated with the vector pointing to each target. All information is relayed to a central computer at Tyndall Air Force Base, where the trajectory of each target, and the missile-drone and missile-interceptor vector distances at the time of simulated warhead detonation, are determined (Figure 2). Two plotting boards display the trajectory data in real time, and the "burst" and "escape" distances are automatically printed out on a typewriter.



Angle-Measuring Equipment

Cubic angle-measuring equipment (AME) performs the basic measurement at each tracking site. The AME operates on the following physical principle: the phase difference between signals received on a pair of spaced antennas is proportional to a direction cosine (angle) of the straight line joining the signal source to the center of the antenna baseline (when the baseline length, d, is much shorter than the slant range, R). The geometry is illustrated in Figure 3. A signal of frequency f_c is radiated from one of the airborne targets, at a distance R from the center of the baseline. The signal received at antenna A has traveled the distance a, and the signal received at antenna B has traveled the distance b. It follows that the phase difference (φ) measured between the antennas is expressed, in wavelengths, as:

$$\varphi = n\lambda = a - b \qquad (1)$$

where λ is the carrier wavelength.

Now, if R >> d, then

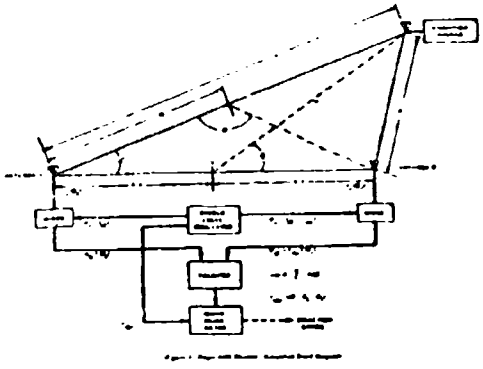
$$\theta \cong \theta', \alpha \cong 90^\circ, \text{ and}$$

$$\varphi = d \cos \theta \qquad (2)$$

where θ is a direction angle of the line joining the signal source to the center of the antenna baseline. The direction cosine, and hence the direction angle, can therefore be determined directly, since the phase difference can be measured and the baseline length is known:

$$\cos \xi = \frac{\varphi}{d} = k\varphi \qquad (3)$$

The electronic equipment required for instrumenting a single AME channel is also indicated in Figure 3. A Double Local Oscillator (DLO) produces three output signals: two (1000 cycles apart) are supplied to the first mixers, and the third is a 1000-cycle reference signal for the phasemeter.



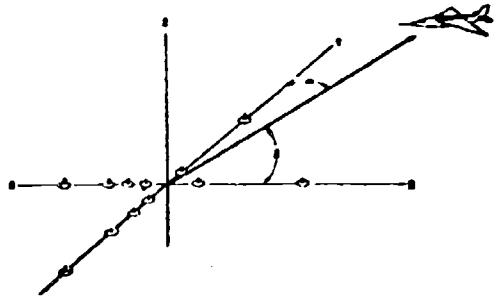
The first mixer associated with each antenna heterodynes the receiver and DLO signals. The resultant i-f signals are 1000 cycles apart, and have the same phase relationship as did the r-f signals received at the antennas. The two i-f signals are fed from the first mixers to the receiver where they are added, amplified, and then detected. The output of the detector is a 1000-cycle signal whose phase is directly related to the phase difference between the two received signals. This 1000-cycle signal, called the data signal, is applied to the phasemeter.

The phasemeter consists of a servo control amplifier and a precision gear train. The amplifier produces gear-train shaft rotation proportional to the phase difference between the data signal (from the Receiver) and the reference signal (from the DLO). Digital encoders, connected to the gear train, digitize the direction cosine data.

Phase-measuring techniques have been developed by Cubic for determining small phase differences with very high accuracies, to the point where measurements exhibit standard deviations in cosine of the order of 15 ppm. In order to obtain such precision, the antenna baseline is 50 wavelengths long (e.g., about 220 ft at 225 mc). In a 50-wavelength baseline, however, phase measurements limited to $\pm$ 180° are identical for any one of 100 different direction cosines between -1 and +1. To eliminate these ambiguities, the baseline has a second pair of antennas 5 wavelengths apart, and a third pair effectively 1/2 wavelength apart. The third pair provides unambiguous data which resolves the 10 ambiguities of the second pair; these in turn resolve the 100 ambiguities of the first pair.

In practice, a "coarse" measurement of the direction cosine is made first, using the third antenna pair. Then, the phasemeter switches to the second ("intermediate") pair to make a somewhat more accurate measurement. Finally, it switches to the 50-wavelength ("fine") pair for the ultimate cosine measurement.

The preceding section has described the equipment associated with one AME and its three pairs of antennas. Two such AME's (six antenna pairs) arranged with mutually perpendicular and bisecting baselines measures two independent direction cosines (ℓ, m) uniquely determining the direction to the target, as shown in Figure 4.



Since MATTS operates on phase-comparison principles, phase distortion within the system must be held to an absolute minimum. Advanced feedback techniques, regulation of critical voltages, precision design of mechanical components, and careful selection of electronic circuits and components therefore must be used throughout the equipment.

Multiple Target Tracking

Multiple-target tracking is achieved by having each of the targets transmit at a slightly different frequency - 221 mc (interceptor), 224 mc (drone), and 225 mc (missile). Common antennas r-f amplifiers and DLO signals, are used for all of the targets, thus producing a different i-f output for each target. Individual i-f amplifiers separate these outputs in the receiver.

From this point on, each target's signal is handled separately as previously described for the single-target system. Compensation for the effectively different baselines (in wavelength units) corresponding to the different received frequencies occurs in the data-reduction process.

Determining Spatial Position

Position in space is determined from the measured direction cosines by triangulations. Figure 5 illustrates the system geometry. The source S is at some location in space described by the Cartesian coordinates x, y, and z. If the direction cosines measured at site #1 (at 0,0,0) are referred to as ℓ_1 and m_1, and the direction cosines measured at site #2 (at 0,H,0) as ℓ_2 and m_2, then:

$$x = R_1 \ell_1 \tag{4}$$

$$y = R_1 m_1 \tag{5}$$

$$z = R_1 \sqrt{1-\ell_1^2 - m_1^2} \tag{6}$$

where:

$$R_1 = \frac{H\ell_2}{\ell_2 m_1 - \ell_1 m_2} \tag{7}$$

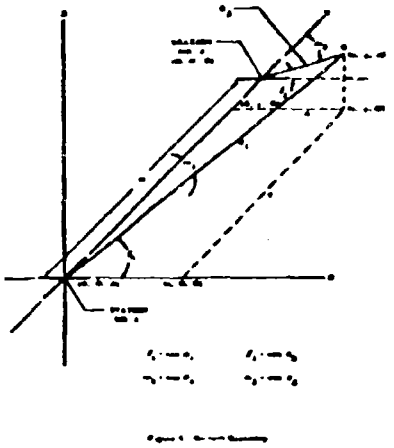
Substituting this value of R_1 into the previous equations yields:

$$x = \frac{H\ell_2\ell_1}{\ell_2 m_1 - \ell_1 m_2} \tag{8}$$

$$y = \frac{H\ell_2 m_1}{\ell_2 m_1 - \ell_1 m_2} \tag{9}$$

$$z = \frac{H\ell_2\sqrt{1-\ell_1^2 - m_1^2}}{\ell_2 m_1 - \ell_1 m_2} \tag{10}$$

Equations (8), (9) and (10) express the Cartesian coordinates of the source position in terms of (1) the direction cosines measured from two separate sites, and (2) the site separation.



Data Handling

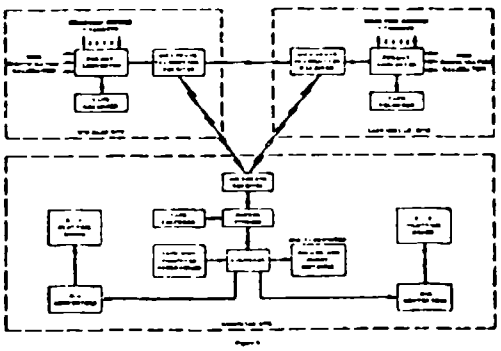
In order to make full use of the system's inherent tracking accuracy, the MATTS output data is processed to a resolution of 5 parts per million. Therefore, each direction cosine is derived as an 1^-bit binary word. An additional bit is required for algebraic sign, since the direction-cosine value may be positive or negative. An Adcon RD-13 optical encoder provides a 13-bit "fine" data word for each direction cosine. A simpler electromechanical encoder (Librascope 709) is coupled through a gear-down arrangement to each Adcon encoder and produces six bits of "coarse" data plus one additional bit for backlash correction.

The encoders are selectively interrogated 20 times, 10 times, or once per second as commanded by MATTS internal timing. Each encoder generates a Gray code number on interrogation, and separate Gray-to-binary conversions are made for the numbers in each encoder pair. Backlash effects are eliminated by combining each pair of "coarse" and "fine" encoder readings to form the final 19-bit word. Each sample comprises six words, corresponding to a pair of direction cosines for each of the three targets.

Calibration numbers representing the variation of each target frequency from an assumed value are derived in each sample interval. The assumed value, in each case, represents a nominal offset between the target frequency and the actual antenna spacing. A burst of 20-kc modulation is inserted in the interceptor's transmission at the instant of missile launching; a similar burst is inserted in the missile's transmission at the instant of simulated warhead detonation.

The six direction cosines and three calibration words, derived at each site during each interrogation, are converted into a special 8-line format, magnetically recorded, and transmitted via r-f link to the computer at Tyndall Air Force Base. The format includes timing data, identification data, and spaces reserved for indicating release and burst, in addition to calibration numbers and direction cosines. Data is transmitted alternately from one tracking site and then the other so as to appear serially at the computer center and thereby minimize buffer storage requirements.

To completely synchronize the data taking and transmission at the two tracking sites, master timing (3 kc) is generated at the Cape San Blas site and transmitted to Carrabelle. There it is employed for synchronizing all pertinent data-handling operations, including interrogating the encoders, synchronizing the format gating operation, and transmitting the format data to the computer. A 1-pps signal is also transmitted from Cape San Blas to provide reset capabilities in the event of momentary loss of the basic 3-kc synchronizing signal or other temporary equipment malfunction.



An input buffer unit at the computer site receives and stores the 12 direction-cosine and six calibration words transmitted in each sample interval from the two stations. (Figure 6). A Bendix G-15D computer is employed for the computation, and one of its memory lines is programmed to synchronize the entry of the data stored in the buffer unit directly into the proper computer memory channels for subsequent operations.

Two separate main programs are stored on an auxiliary tape unit associated with the computer. These program (1) first-mode operation, where trajectory points are computed, and (2) second-mode operation, where burst and escape vector calculations are performed. During the trajectory computation, interceptor and drone paths are computed at the rate of one x-y-z set per second. The x and y coordinates of interceptor and drone are simultaneously plotted on the first plotting board, and the corresponding y and z values are plotted on the second.

After the missile launch signal is received, the plotting boards display missile and drone trajectories. On receipt of the burst signal, the second ("burst") program on the auxiliary magnetic tape unit is fed into the computer. The computer computes the drone and missile x-y-z at simulated burst, and from this information determines the vector miss distance. The computer also computes the interceptor x-y-z at simulated burst, and from this information determines the vector escape distance. These values are printed out by an electric typewriter.

In the "burst" program, linear interpolation is employed between the x, y and z values from the samples immediately before and after to determine exact positions at the time of simulated burst. During this final computation, the computer also computes the interceptor trajectory, so that the pilot's escape maneuver may be evaluated from inspection of the plotting board.

The MATTS data-handling equipment includes a number of self-checking features. First of all, known Gray code numbers may be inserted at the input, and the final output format values checked against a stored, correct value. Any difference between the two is externally signaled to indicate equipment malfunction. Marginal checking is also provided where the known Gray coded number is cyclically inserted into the input and certain voltages are increased and decreased in known sequence in various parts of the equipment, with malfunctions again being observed. Defective circuit boards may be readily located by this technique and replaced, since the

voltage variations are confined to particular divisions of the equipment.

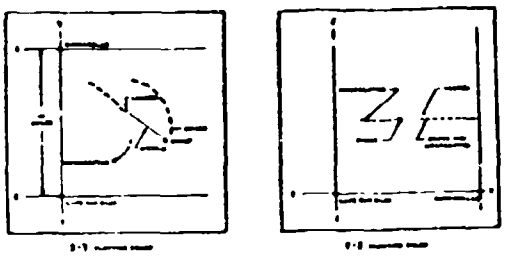


Figure 7 shows a typical intercept mission plotted on the x-y and y-z boards. The plotted drone track continues to point of burst, while the interceptor path is indicated up to the time of launch. The missile track is plotted from launch to burst, while the interceptor track is shown following burst. In the particular mission shown, the drone and interceptor are at substantially the same altitude, with the resultant superposition of their tracks on the y-z board.

The employment of this AN/GSG-29 (XI-1) two-station tracking system for evaluation of intercept missions, as described, represents only one of its capabilities. Other applications of similar systems include range safety, drone guidance, missile evaluation, bomb scoring, air traffic control, simulated firings, and formation attack scoring.

PRACTICAL STRIP TRANSMISSION LINE COMPONENT DESIGN

By

V. T. Norwood
Engineering Division
Hughes Aircraft Company
Culver City, California

The analysis of strip transmission line was covered thoroughly and reported in symposia five years ago. Cohn, who was responsible for most of the original analysis, has recently reviewed the status of strip transmission line*, and has assembled an excellent bibliography of 65 papers on the subject.

The Need for Strip Transmission Line and Choice of Line Type

Our original motivation for employing strip transmission line in receiver designs was the need for miniaturization. After becoming involved in component design work, the additional advantages of production control and accuracy have made the method attractive even where packaging requirements are not stringent.

Initially, components which were offered as a "shelf-item" by various manufacturers were evaluated and found unsuitable for our very sensitive receiver applications. We decided to design components so that performance could be optimized in the frequency ranges required and so that each circuit could be shaped for the available space. Strip transmission line is best exploited by employing meticulous design techniques and by choosing circuit layout to conform to the system requirements in every detail.

The first applications for which strip transmission line offered the only solution was one involving mixers supplied from a common local oscillator line. The mixers with their i-f matching circuits and filters were to weigh less than four ounces and were to be positioned on the rear of a tracking antenna dish in a thin annular region. In addition, r-f leakage and accoustical noise vibration requirements were severe. With these objectives in mind, the transmission line cross-section shown in Figure 1 was evolved.

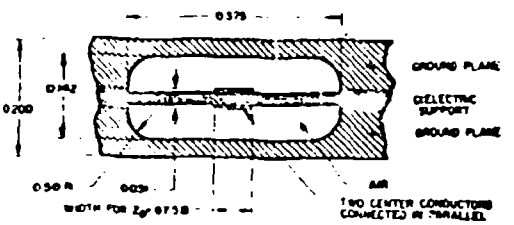


Figure 1 Cross Section of the Strip Transmission Line.

Two ground planes are clamped together with the dielectric sheet in between, supporting the conducting strips. For a sheet containing an entire r-f circuit with many components, the ground planes are fastened along each strip edge so that the finished effect is that of several tunnels as shown in the Figure 2.

This is a typical circuit which was photo-etched from doubly copper-clad KEL-F sheets. The desired inner conductor pattern is first inked very precisely in as large a scale as is practical (a factor of ten is often necessary). The drawing is then photographed in the proper scale for etching directly. For production items, a control pattern should be included on every photograph. A ground plane drawing is made with reference holes to coincide with those on the inner conductor drawing. If the circuit is to be fabricated in large quantities, the sloped sides are drawn and photo-milling is used. Laboratory models are generally made with straight sides by end mills and if a complicated set of ground planes is to be duplicated as many as eight to ten times, a template for a reproducing mill is made.

The curved ground planes afford a rigid mechanical structure with accurate, ground plane positioning. A mixer in this form has functioned within specifications during shock and vibration tests. This configuration also compensates for impedance variation caused by non-uniformities in the dielectric supporting sheet since the ground plane separation increases with sheet thickness. The outermost edges of the circuit can be protected against external radiation by incorporating a lip on either ground plane. Internal radiations are curtailed by good r-f matching at every junction.

Before adopting transmission line dimensions, it was necessary to determine the effect of curved ground planes upon the electric field of the conducted wave. Although the effect of vertical walls can be deduced from the transverse attenuation expression (27 db of attenuation for each unit of lateral displacement equivalent to the ground plane spacing), it is still instructive to measure the impedance for a given strip configuration by means of a d-c resistance analogue method.

* "A Reappraisal of Strip Transmission Line", Seymour B. Cohn, "The Microwave Journal", Volume 3, No. 3, March, 1960.

The conducting elements of a cross-section of strip line are painted on Teledeltos paper with as large a scale factor as can be accommodated. A calibrating figure of two coaxial rings is drawn on the same paper and the d-c conductivity between inner and outer conductors for the two figures are compared. It is important to use circular symmetry in the calibrating figure, and it is advisable to check the unknown figure in two directions to resolve any grain effect in the paper. Conductivity measurements will also vary with humidity, so that both measurements must be made at the same time. It was decided to keep the vertical portion of the ground planes 1.5 strip widths away from the widest strip likely to be used so that the impedance curves* for a strip between two infinite slabs would apply.

The ground plane spacing was then determined by calculating the smallest dimensions which permitted a 130 ohm strip to be etched within the photographic tolerances of one mil. Strip thicknesses were limited to those in which dielectric sheets can be obtained with good quality control. Our original choice was influenced by sheet thicknesses which were standard; however, subsequently, our own plastics experts have found it necessary to develop methods for producing copper clad sheets with good adhesion properties and thickness variations no greater than ±0.001. At first it was thought that the ground plane spacing should be large enough to accommodate the standard r-f diode package, but it was foreseen that smaller diodes would become available and this consideration was discarded. Large transmission line dimensions have been used for special applications such as varactor diode circuits in which miniaturization must be compromised in order to provide high Q; however, the circuits discussed here have all been made with the 0.14" ground plane spacing shown in the sketch.

Measuring Methods

A lamentable amount of effort was expended in designing and building movable probes in strip transmission line. Both vertical and horizontal probes and carriages have been tried and found troublesome. Probe alignment is tedious and must be checked constantly. One must have simple joints in the inner conductor in order to have a versatile traveling probe and this has proved difficult in our configuration.

The method which has proved most valuable to us involves the use of a good commercial slotted line with an accurate, repeatable slide screw tuner. Any equipment between the line and the component under test may then be nullified.

In the beginning one is in the awkward position of needing a good load in order to design a transition, and a good transition to design a load. This pair can be refined very quickly after a few tries by overdesigning the load taper. A series of good in-line transitions was first designed, having VSWR's of 1.02 in narrow frequency bands. Transitions for actual components were then made covering 500 mcps with VSWR < 1.05. With a well matched transition and a movable short having a VSWR of 30, the wavelength in a 65 ohm strip was measured. A constant value for λ_g of .925 λ_0 was measured at 3 kmcps, 6 kmcps, 9 kmcps and 12 kmcps.

Verification of this measurement was obtained when a junction was formed by removing the dielectric on either side of a conducting strip. The resulting mismatch caused a VSWR of 1.08, whereas the value calculated from the λ_g/λ_0 ratio is 1.075.

In the design of components for laboratory testing, machine shop capabilities are often influential. For example, a short curved to accommodate an end mill was found to be just as effective as a square short; therefore, quarter wave stubs which occur in strip transmission line designs now have semi-circular ends.

Components

Following are described components which have been designed for use in many different circuits.

Proximity couplers were designed in the form shown in Figure 3. The dielectric was cut away from the coupling region in order to avoid the effects of dielectric inhomogeneities on the closely confined antisymmetrical mode. The coupling variation with strip separation is linear for couplings from 10 db to 40 db and has a slope of 189 db per inch. This corresponds very closely with the expression $e_t \cdot 27/b$ where e_t is the transverse attenuation in db per unit length, and b is the ground plane spacing. This coupling relationship and constant directivity of 8 db holds throughout the frequency range from 8.5 kmcps to 12 kmcps.

The 8.5 directivity is significantly less than the 20 db which would be expected from theory. The directivity can be improved by using matching posts, but there is an attendent diminution in bandwidth. Equipment used for obtaining measurements of the coupling and directivity appears in the referenced figure.

* Seymour Cohn, "Characteristic Impedance of Shielded Strip Transmission Line", Volume 2, IRE-MTT No. 2, July 1954, pp. 52-27.

A double coupler was designed using the same dimension shown for the single coupler. With the unused arms matched, the two couplers were completely independent in coupling and directivity and each evidenced the characteristics of a single coupler. Mismatching various arms caused power to be reflected into other arms in a completely predictable manner. The mitre bends used in the couplers caused a VSWR of 1.09.

The conventional 3/2 λ hybrid ring has been found to exhibit excellent power division and VSWR's less than 1.1 over the frequency band from 8.7 kmcps to 9.4 kmcps. The insertion loss from one arm to another is about 0.15 db.

Holders for 1N23E and 1N415D crystals have been designed using coaxial sections with transformers for adapting from strip transmission line to the crystal shell. In each case, the noise figure for many samples has met those stated by the manufacturer for standard waveguide mounts. A major cause of difficulty in designing good, rugged miniature crystal mounts is the variation caused by torsional strain on the diode contacts. Good mechanical support must be provided without gripping the terminals so firmly that vibrations are translated directly to the diode junction.

Transitions from strip transmission line to waveguide have been designed according to the sketch shown in Figure 4. In one case, the VSWR did not exceed 1.1 from 9.5 kmcps to 11.0 kmcps and in the other case, the same performance extended from 8.5 to 9.6 kmcps.

A perpendicular transition from miniature coaxial line to strip transmission line was designed and the VSWR was less than 1.1 from 8.6 kmcps to 9.6 kmcps.

Capacitive filters for passing microwaves and inhibiting intermediate and lower frequencies have been designed using the "dove-tail" form seen in the circuit photograph. The VSWR does not exceed 1.1 from 8.75 kmcps to 9.3 kmcps. It was found necessary to use gaps of four mils and to draw the artboard in times fifty scale. It was determined experimentally that an etching tolerance of ±0.001 could be used.

The three-way power divider shown in Figure 1 is fed with miniature coaxial cable which is perpendicular to the stripboard. In the first application (not shown), the permissible phase discrepancy among the three paths was less than 1° or 1 mil per inch. We thought it important to choose a symmetrical junction even though a perpendicular type is often hard to match in a small space. It was not possible to retain symmetry in the lines themselves so the final lengths were chosen experimentally. The lengths were first drawn with equal center line lengths and these were found to differ by 30 degrees in three inches. The VSWR on the input arm was less than 1.4 in the frequency range from 8.8 kmcps to 9.4 kmcps.

The photograph shows the positioning of i-f matching elements on the strip transmission line. This utilizes space otherwise wasted and permits short line lengths. Where it is possible, the i-f input coils might also be included on the r-f circuit.

Conclusions

In designing a strip transmission line circuit, one should arrange his components so that the maximum number of components may be placed on one board, thus insuring exact reproduction of line lengths, minimizing manufacturing costs, minimizing connectors and optimizing space usage. A single strip transmission line component, such as a mixer, does not compare favorably in size with its waveguide counterpart because the transitions must be included in the competition; however, as more components are added, the waveguide circuit grows relatively more cumbersome.

The success of the final design rests on the drafting technique and the fabrication control. The latter includes careful monitoring of the photographic process and use of materials that are dimensionally stable with good strip adhesion. Capable plastics advisors can offer invaluable assistance to the electrical designer in getting the most out of strip transmission line for each application.

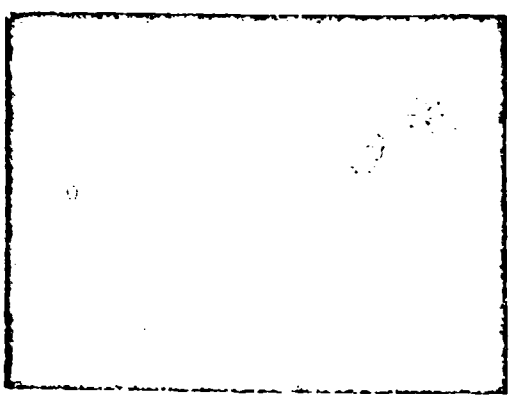


Figure 2 Typical Receiver Circuit

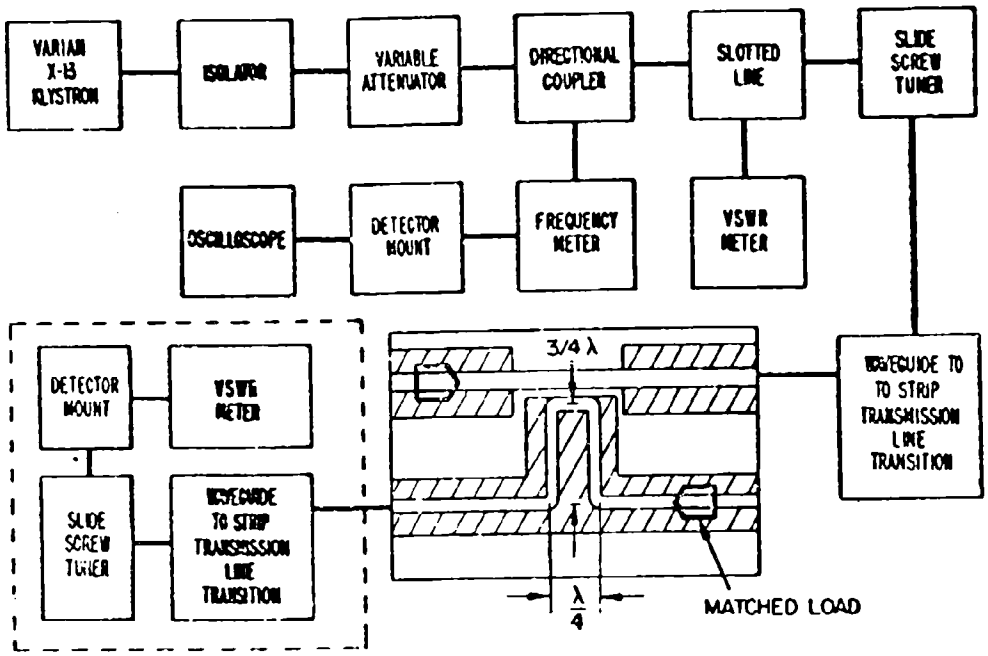


Figure 3 Test Set-up for Coupler Measurements

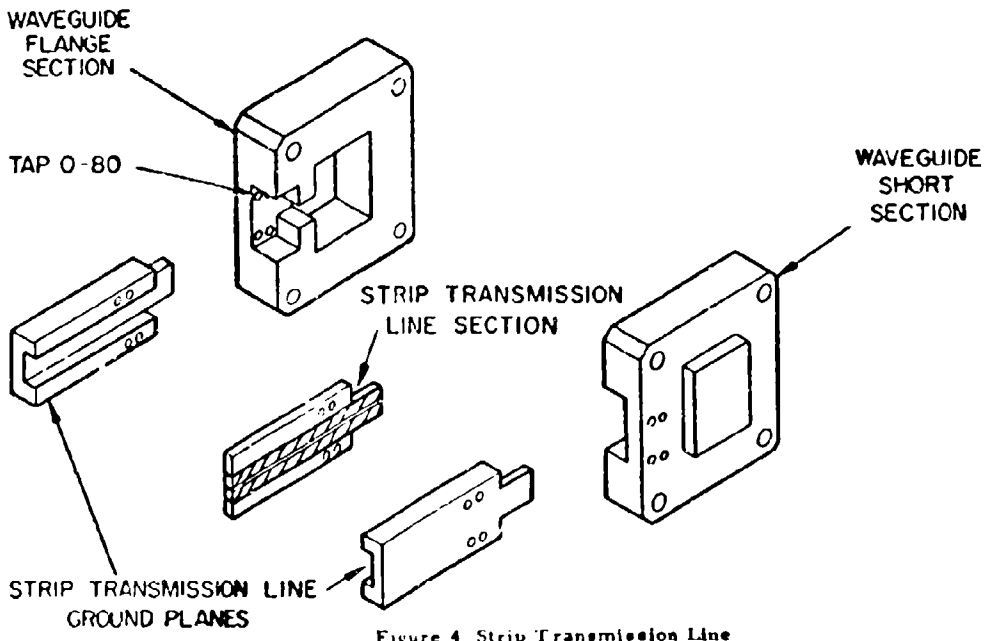


Figure 4 Strip Transmission Line to Waveguide Transition

A NEW GYRO FOR AUTOPILOT USE

By: Sahag Dardarian, Project Engineer, Kearfott Division, General Precision, Inc.

INTRODUCTION

This paper discusses a single-degree-of-freedom rate gyro particularly suited for autopilot use. This unique gyro senses and integrates rate by means of a capacitor. The output is a d-c signal proportional to the sum of the attitude and rate.

Typical applications of the Capacitor Rate Integrating Gyro (CRIG), may be for stabilization or telemetering where a rate signal, or a combined attitude and rate signal is required. This small, rugged, lightweight gyro is ideally suited for autopilots in target drones, short-range missiles, small aircraft and boats.

The instrument may be used as an azimuth gyro monitored by a magnetic compass. Or, it may be employed to hold verticality by means of a bubble-switch control. Scale factors are readily changed by varying capacitance and resistance. The gyro is easily zeroed by simply short-circuiting the capacitor.

PRINCIPLE OF OPERATION

This paper presents two methods of rate integration, both using a gyro as a sensor. It also presents some typical values which can be expected from such a device. Close investigation will reveal that both methods are analogous. The method seen in Figure 1 is a mechanical system analogous to CRIG. Both are described below.

In the mechanical analog (See Figure 1), an input rate $\dot{\theta}$ produces a torque around the precession axis. This torque is opposed by a spring-lever-dashpot system. If we assume the dashpot piston is welded to the cylinder walls, the gyro would then operate as a conventional rate gyro, where the lever deflection is proportional to the rate input. If the dashpot piston were free, it would be moved by a reaction force of the stretched spring. In an ideal dashpot, the force on the piston is proportional to the piston velocity. With a constant-rate input, the piston would move with uniform velocity inside the cylinder. The Lever L, through which the torque is applied,

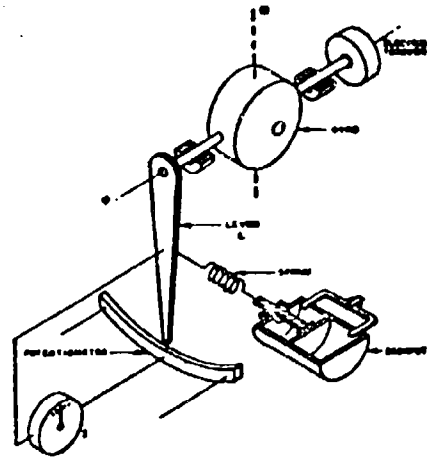


Figure 1 - MECHANICAL ANALOG OF THE ELECTRIC CRIG

rotates to maintain the dynamic torque equilibrium. The wiper on the lever arm measures the total precession angle, ψ. This is equal to some constant times the rate input (due to the initial extension or compression of the spring) plus another constant times the total time integral of the rate (due to the displacement of the piston).

To put command signals into the gyro, an electrical torquer can be placed on the precession axis to simulate input torques to the gyro. The unbalanced command torque will displace the gyro causing a deviation from null position. The resulting output signal will appear as an error signal in the autopilot. The signal will be nulled out by a real input rate from the airframe.

Figure 2 illustrates the electrical CRIG. It consists of a single-degree-of-freedom gyro, a pick-off potentiometer and a voice-coil torquer.

Input rates, $\dot{\theta}$, precess the gyro around the ψ axis. Each of the wipers on the moving armature sweep across the pick-off potentiometers. The moving armature is connected to the gyro motor housing. The voltage gradient of each potentiometer is in direct ratio and opposite in sign. As

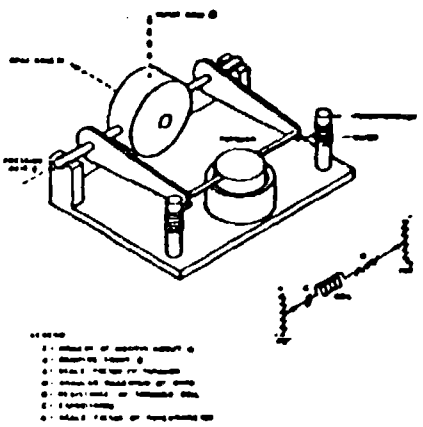


Figure 2 - ELECTRICAL CRIG

a result, when the gyro precesses a potential exists between the wipers. This potential is placed across the circuit parameters seen in Figure 2. If we neglect the capacitor (consider it shorted), the current would flow through the torquer coil only. The instrument would behave as a rate gyro; that is, the current in the torquer coil is proportional to the rate input $\dot{\theta}$.

If the short across the capacitor were removed and the gyro put in operation, the capacitor would store the charge that flowed thru the torquer coil. As this occurred, the voltage across the capacitor would increase and that across the torquer coil would decrease. This results in a smaller current flow. If this current proved inadequate to restrain the gyro, the gyro would precess to a new position capable of satisfying the torque requirement.

The voltage output from the potentiometer is the sum of the voltage on the capacitor. This is the time integral of the current into the torquer coil plus the voltage across it. The latter voltage is proportional to the rate input. Thus the gyro voltage output is proportional to rate plus attitude.

$$\text{or} \qquad Ey = k\dot{\theta} + k'\theta$$

LIST OF SYMBOLS

C	Capacitance
E	Supply voltage
H	Angular momentum of gyro-wheel
I_p	Moment of inertia around precession axis
L	Inductance of torquer coil
R	Resistance of torquer coil
S	Laplacian operator
i	Instantaneous torquer current
t	Time
y	Ratio of wiper displacement to maximum displacement
θ	Input angle
λ	Torquer scale factor
ψ_m	Maximum angular displacement
$\omega_u\dot{\theta}$	Input angular velocity

Analysis of Gyro Equations:

Summing torques around the precession axis:

$$I_p \psi_m \ddot{y} - H\dot{\theta} + \lambda i = 0$$

Summing voltages around the loop neglecting phase shift:

$$L\,di/dt + Ri + 1/C \int i\,dt = yE$$

$$\text{or} \quad \frac{dy}{dt} E/L = \frac{i}{L} C + \frac{R}{L} di/dt + \frac{d^2 i}{dt^2}$$

In LaPlace notation we may rewrite the above as:

$$I_p \psi_m S^2 y + \lambda i = HS\theta$$

$$\text{and} \quad -\frac{E}{L} Sy + (S^2 + R/L\, S + 1/LC)i = 0$$

The determinate of the set is:

$$D=\begin{vmatrix} I_\rho \Psi_m S^2 + & \lambda \\ -\frac{E}{L}S & S^2+\frac{R}{L}+\frac{1}{LC} \end{vmatrix}$$

$$D=S\begin{vmatrix} I_\rho \Psi_m S & \lambda \\ -E/L & S+R/L+1/LC \end{vmatrix}$$

$$D=S\left[I_\rho \Psi_m S\left(S^2+\frac{R}{L}S+\frac{1}{LC}\right)+\frac{\lambda E}{L}\right]$$

$$\theta_o/y=\frac{\begin{vmatrix} HS & \lambda \\ 0 & S^2+R/LS+1/LC \end{vmatrix}}{D}$$

$$\frac{HS\left(S^2+\frac{R}{L}S+\frac{1}{C}\right)}{D}$$

$$\theta_o/y=\frac{H\left(S^2+R/LS+1/LC\right)}{I_\rho \Psi_m S\left(S^2+R/LS+\frac{1}{LC}\right)+\frac{\lambda E}{L}}$$

$$\theta_o/y=\frac{H\left(LCS^2+RCS+1\right)}{I_\rho \Psi_m S\left(LCS^2+RCS+1\right)+\lambda EC}$$

Let $\theta_o=W/S^2$

$$\boxed{W/y=\frac{H\left(LCS^2+RCS+1\right)}{\left[I_\rho \Psi_m S\left(LCS^2+RCS+1\right)+\lambda EC\right]S^2}}$$

Similarly, one may solve for i/θ_o and i/ω and find,

$$\frac{i}{\theta_o}=\frac{EHS}{L\left[I_\rho \Psi_m S\left(S^2+\frac{R}{L}S+\frac{1}{LC}\right)+\frac{\lambda E}{L}\right]}$$

$$\frac{i}{\omega}=\frac{EH}{LS\left[I_\rho \Psi_m S\left(S^2+\frac{R}{L}S+\frac{1}{LC}\right)+\frac{\lambda E}{L}\right]}$$

Assuming L is small:

$$y/\theta_o=\frac{H\left[RCS+1\right]}{I_\rho \Psi_m S\left[RCS+1\right]+\lambda EC}$$

or:

$$\frac{\lambda EC}{H}\,y/\theta_o=\frac{RCS+1}{\frac{I_\rho \Psi_m RC}{\lambda EC}S^2+\frac{I_\rho \Psi_m}{\lambda EC}S+1}$$

$$\frac{y}{\theta_o}=\frac{H}{\lambda EC}\left[\frac{as+1}{\frac{S^2}{\omega_i^2}+\frac{2z}{\omega_i}S+1}\right]$$

where $RC=a \qquad \omega_i^2=\frac{\lambda E}{I_\rho \Psi_m R} \qquad \frac{2z}{\omega}=\frac{I_\rho \Psi_m}{\lambda EC}$

$$z=\frac{1}{2RC}$$

$$y/\theta_o=\frac{H}{\lambda EC}\left[1+\frac{1}{1-z^2}(1-a z\omega_i)^2+\omega_i^2(1-z^2)(a)^2\right]^{1/2} e^{-z\omega t}\sin\left(\omega_i\sqrt{1-z^2}\,t+\Psi\right)$$

or:

$$y/\theta_o=K\left[P+Ue^{bt}\sin\left(Rt+\gamma\right)\right],$$

K, P, U, b, R, and γ are constants

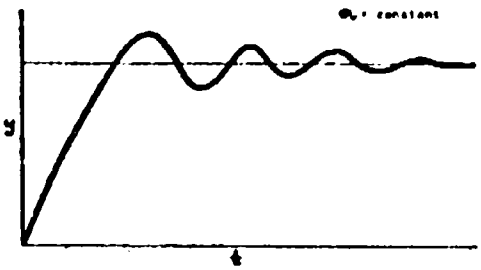


Figure 3

Figure 3 is the CRIG output versus time for a step input, Θ_o. The voltage rises, overshoots and then steadies at the value yE. If, due to the resolution of the potentiometer, this value can not be obtained, the gyro will continue to hunt, as a "bang-bang" servo, around the steady-state value yE. The time average, however, will be accurate.

If we again assume L is small:

$$\frac{\gamma}{\omega} = \frac{H(RCS+1)}{S^2\left[\Psi_m I_\rho S(RCS+1)+\lambda EC\right]}$$

$$\frac{\gamma}{\omega} = \left[\frac{H}{\lambda EC}\right]\frac{RCS+1}{S^2\left[\frac{I_\rho \Psi_m S}{\lambda EC}(RCS+\beta+1)\right]}$$

$$\frac{\gamma}{\omega} = \frac{H}{\lambda EC}\left[\frac{RCS+1}{S^2\left[\frac{I_\rho \Psi_m RC}{\lambda EC}S^2+\frac{I_\rho \Psi_m S}{\lambda EC}+1\right]}\right]$$

$$\frac{\gamma}{\omega} = \frac{H}{\lambda EC}\left[\frac{aS+1}{S^2\left[\frac{S^2}{\omega_1^2}+\frac{2z}{\omega_1}S+1\right]}\right]$$

where $a = RC$

$$\omega_1 = \frac{\lambda E}{I_\rho \Psi_m R} \qquad \frac{2z}{\omega_1} = \frac{I_\rho \Psi_m}{\lambda EC}$$

$$\therefore 2z = \frac{1}{RC} \qquad z = \frac{1}{2RC}$$

$$\frac{\gamma}{\omega} = \frac{H}{\lambda EC}\left\{t + a - \frac{2z}{\omega_1} + \frac{\left[(1-az\omega_1)^2+\omega_1^2(1-z^2)(a)^2\right]^{1/2}}{\omega_1\sqrt{1-z^2}}\right.$$

$$\left. e^{-z\omega t}\sin(\omega\sqrt{1-z^2}\,t-\psi)\right\}$$

$$\frac{\gamma}{\omega} = \frac{H}{\lambda EC}\left[t+P+Ke^{bt}\sin(Rt+\lambda)\right]$$

where H, λ, E, C, P, K, b, R, and γ are constants

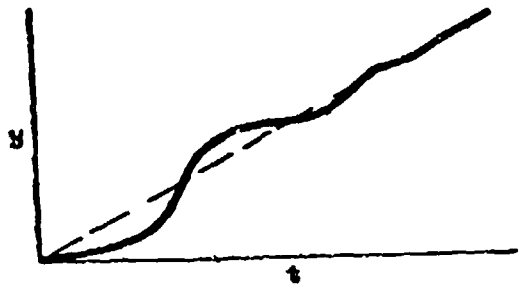


Figure 4

$$\frac{d}{dt}\frac{\gamma}{\omega} = \frac{\dot{\gamma}}{\omega} = \frac{H}{\lambda EC}\left[1+Te^{bt}\sin(At+\mu)\right]$$

Figure 4 shows the gyro output for a step ω input. At first, the current lags the true value (shown dotted). The current soon reaches the true value, overshoots and finally damps out to the true value. This occurs when the armature is moving with constant angular velocity. That is, when the inertia torque term reaches zero.

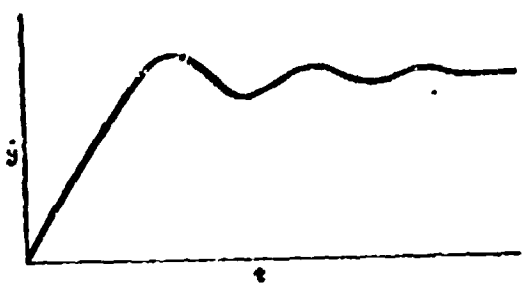


Figure 5

Figure 5 shows the derivation of $\dot{\gamma}$ versus time. The steady-state value of $\dot{\gamma}$ is seen as a constant for a constant ω input.

Again assuming L is small:

$$\frac{i}{\beta_0} = \frac{EHS}{\left[I_\rho \Psi_m S(Rs+1/C+\lambda E)\right]}$$

$$\frac{i}{\beta_0} = \frac{CHES}{I_\rho \Psi_m CRs^2 + I_\rho \Psi_m s + \lambda EC}$$

$$\frac{i}{\theta_o} = \frac{H}{\lambda}\left[\frac{S}{\left(\frac{I\rho\Psi_m R}{\lambda E}S^2 + \frac{I\rho\Psi_m}{\lambda EC}S + 1\right)}\right]$$

$$\frac{i}{\theta_o} = \frac{H}{\lambda}\left[\frac{S}{\frac{1}{\omega_1^2}S^2 + \frac{2z}{\omega_1}S + 1}\right]$$

$$\omega_1^2 = \frac{\lambda E}{\Psi_m I\rho R} \qquad \frac{2z}{\omega_1} = \frac{I\rho\Psi_m}{\lambda EC}$$

$$\frac{i}{\theta_o} = \frac{H}{\lambda}\left[\frac{\omega_1^2 e^{-z\omega_1 t}}{\sqrt{1-z^2}}\sin(\omega_1\sqrt{1-z^2}\,t + \Psi)\right]$$

Assuming a step Θ input:

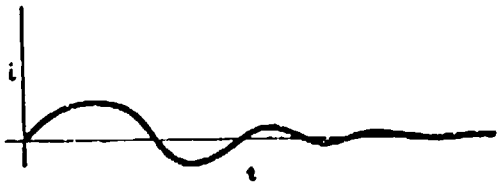


Figure 6

Figure 6 shows that for a step angular input the current i rises to a maximum and oscillates back to zero.

$$\frac{i}{\omega} = EH\left[\frac{1}{S(I\rho\Psi_m Rs^2 + \frac{I\rho\Psi_m}{C}S + \lambda E)}\right]$$

$$\frac{i}{\omega} = \frac{H}{\lambda}\left[\frac{1}{S(\frac{I\rho\Psi_m R}{\lambda E}S^2 + \frac{I\rho\Psi_m}{\lambda EC}S + 1)}\right]$$

$$\frac{i}{\omega} = \frac{H}{\lambda}\left[1 + \frac{1}{\sqrt{1-z^2}}e^{-z\omega t}\sin(\omega_1\sqrt{1-z^2}\,t + \psi)\right]$$



Figure 7

Figure 7 shows the output current versus time for a constant $\dot{\theta}$ input. The current rises to a steady-state value proportional to the input rate.

The steady-state transfer function of $i/\dot{\theta}$ is:

$$\frac{i}{\dot{\theta}} = \frac{H}{\lambda}$$

$$\text{or} \quad i = \frac{H}{\lambda}\dot{\theta}$$

Summing voltages around the loop, the following is true:

$$yE = \frac{1}{C}\int i\,dt + iR$$

$$yE = \frac{1}{C}\int \frac{H}{\lambda}\frac{d\theta}{dt}\,dt + \frac{HR}{\lambda}\dot{\theta}$$

$$yE = \frac{H}{C\lambda}\theta + \frac{HR}{\lambda}\dot{\theta}$$

and therefore at steady state:

$$yE = k\theta + k'\dot{\theta}$$

Performance Characteristics

A. Maximum Rate Range

From the analysis, it can be seen that the natural frequency of the gyro is given by:

$$\omega_\eta = \frac{1}{2\pi}\sqrt{\frac{\lambda E}{\Psi_m\ I\rho\ R}}$$

The resistance R is the total circuit resistance. Figure 8) shows the natural frequency versus external resistance for a typical gyro.

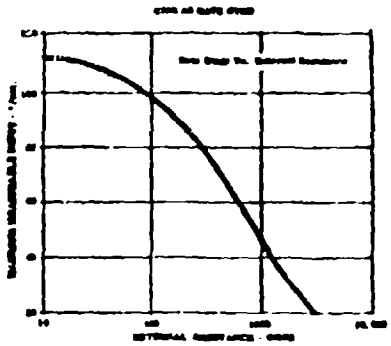


Figure 8a - RATE RANGE VS. EXTERNAL RESISTANCE



Figure 8b - NATURAL FREQUENCY VS. EXTERNAL RESISTANCE

The steady-state transfer function of θ/ω is:

$$\theta/\omega = \frac{H}{\lambda}$$

$$\text{or } i = \frac{H\omega}{\lambda} = k_1\omega \qquad ; \quad k_1 = \frac{H}{\lambda}$$

If a rate is put into the gyro when it is acting as a rate gyro (C is shorted), the voltage output is:

$$E = iR = k_1 R\omega$$

since E is bounded by the voltage on the potentiometers. Then the product of R and ω is a constant. The maximum measurable rate, ω_{max}, is seen in Figure 8a.

B. Maximum Attitude Range

The steady-state voltage output from the gyro with a θ input is:

$$Ey = E_{out} = \frac{H\theta}{\lambda C} = k(1/C)\theta \qquad ; k = \frac{H}{\lambda}$$

The maximum integrable angle θ is dependent upon the value of capacitance C since the maximum E out is fixed. Figure 9 shows the relationship between the maximum attitude angle versus capacitance (assuming $\theta \longrightarrow 0$ when $\theta \longrightarrow \theta_{max}$).

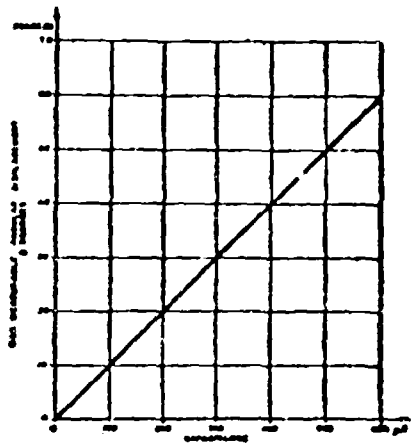


Figure 9 - CRIG AS A RATE INTEGRATING GYRO (STATIC RANGE, $\theta = 0$) ATTITUDE RANGE VS. CAPACITANCE

C. Drift Rate

The drift rate of this gyro, as that of any other integrating gyro, depends upon the built-in restraints. Pigtail (cat whiskers) torques, stray magnetic field torques, mass unbalance torques, bearing friction torques and potentiometer friction torques must be held to an absolute minimum.

To analyze these factors is not the purpose of this paper. However, their composite effects can be held to a value low enough to keep the drift rate to within 1°/minute. The largest frictional torque is that due to the potentiometer. This may be eliminated at the expense of increasing the required electronics.

A simple E-bridge type take-off may be substituted for the potentiometer as an angle transducer. Its output can be fed into an amplifier, demodulated and returned to the torquer coil as a direct current. This current is then the output proportional to rate.

Damping must be provided in this instance as the loop is unstable without the potentiometer friction.

In using the potentiometer-type CRIG, a damping of 0.2 to 0.3 critical (coulomb equivalent) is obtained from the potentiometer friction.

Summary

The CRIG is a compact, rugged, light-weight gyro. It is ideally suited for use in autopilots which require a signal proportional to attitude and rate.

The natural frequency of the instrument could reach 35 cps.

The maximum input rate depends upon the external resistance of the gyro. It may be as high as 110°/sec.

The maximum attitude indication is proportional to $1/c$. Using a 500 μfd capacitor, a range of 50° can be sensed if the rate is negligible when the angle approaches 50°

The drift of the gyro can be held to ±1°/minute. This value can be decreased if an E-bridge type take-off is used instead of the indicated potentiometer.

FREQUENCY SELECTION

By: Mr. H. R. Smith, Signal Communications Department,
U. S. Army Electronic Proving Ground, Fort Huachuca, Arizona

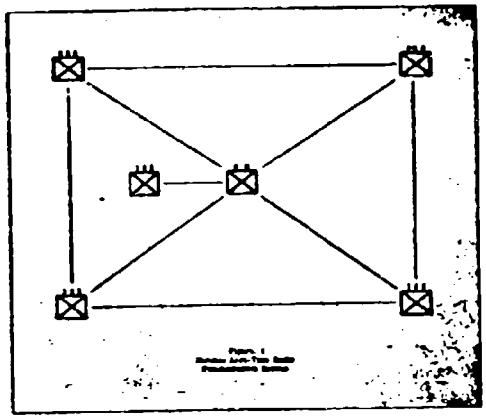
At some future date it is anticipated that signal frequency engineering officers, functioning as part of the command staff, will have at their disposal field type data processing equipment. When such equipment is available, it is our intention to provide a tested and proven radio frequency selection technique utilizing the high speed computational capability of data processing equipment. As this frequency selection method is in the process of development, frequency assignment lists will be generated using approved tables of organization and equipment (TO&E) as the basis for frequency requirements. Additionally, with the co-operation of units already in the field, frequency assignment lists will be generated reflecting the requirements and limitations peculiar to the geographic location and deployment of those units. These lists may then be employed by the user for a two-fold purpose: to familiarize the user with the capabilities and limitations of the process and also to evaluate the adequacy and accuracy of the technique.

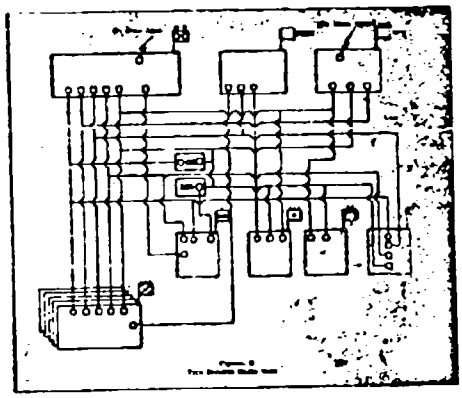
There are several kinds of data required for effective and efficient frequency assignment. For simplicity, categorization of data can be considered in three classes. One, operational data which is based upon frequency needs, equipment types and the doctrine of employment of the radio communication complex is required to be supplied by the user. Two, technical data must be acquired as part of the frequency assignment method, which includes that information required to compute the total electromagnetic environment contributions, in both frequency and magnitude, of all devices under consideration. Three, computational means of applying the technical data to implement the operational requirements in the most efficient and effective manner is necessary.

A general frequency selection method, currently somewhat limited by available data and techniques, has been developed encompassing all three requirements and has been applied to a specific problem.

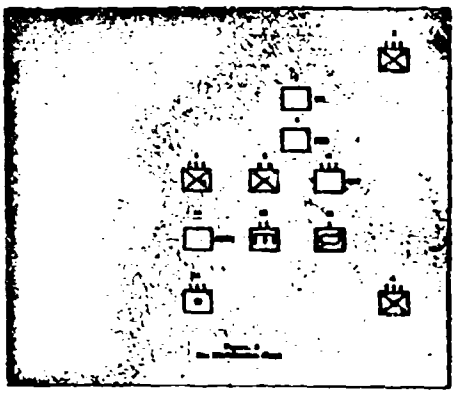
The operational data for this specific problem was derived from a training text prepared for use in training a signal battalion. The radio network structure is displayed in two sections. First, a multi-channel radio relay system is employed to inter-connect radio sites as shown in Figure 1. Second, a structure of push-to-talk radio nets is employed as illustrated in Figure 2. There are a total of fourteen radio sites, eighteen radio nets and eighty-one radio sets used throughout. Of the eight equipment types used, three types are AM and five types are FM. Power output of the various transmitters ranges from 1 watt to 100 watts. Total tuning ranges of these equipments is from 1.5 mc. to 20.0 mc. and from 38.0 mc. to 70.9 mc. with the AM equipment occupying the lower band of frequencies and the FM equipment occupying the higher band.

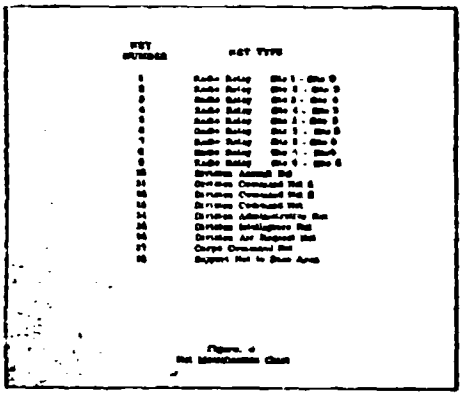
The display of the communication structure in Figures 1 and 2 is rather difficult to visualize. In order to assist in the picturization of these equipment-frequency relationships a method of presentation of this information was developed.

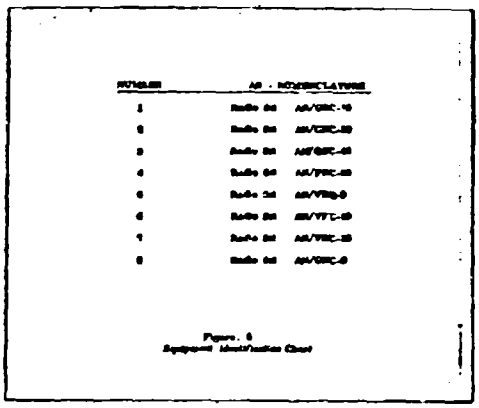




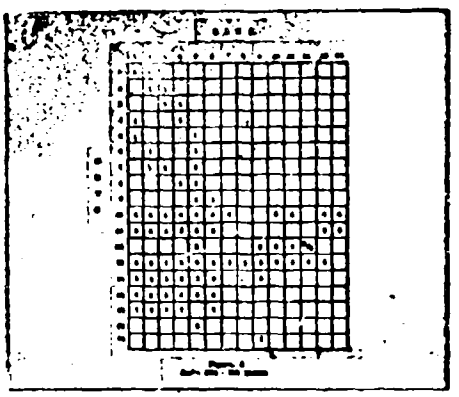
Inspection of Figures 1 and 2 reveals that there are fourteen separate and distinct radio sites. These fourteen sites are arbitrarily spaced and numbered (Figure 3.) in a site identification chart. Considering each link of the radio relay system as a net, there are eighteen nets. These are also arbitrarily numbered (Figure 4.) for identification. There are eight different radio sets, using the "AN" nomenclature, used to implement the communication requirements. These equipments are arbitrarily numbered for identification. (Figure 5.)







Using a matrix, wherein the radio site numbers are plotted against radio set numbers, it is possible to show the network structure and equipment used at each site in each net in a form amenable for ready inspection. (Figure 6.) At each net-site intersection in the matrix where a radio set is required, the number of the radio set is placed. The maximum possible size of the matrix is essentially unlimited. However, the matrix illustrated here is only of the size necessary for this specific problem.



The site-net matrix also points up equipment density distribution. At site 5 (Division Main) there are twelve radio sets while at site 8 (Division Signal Officer) there is only one. It is also interesting to note that at site 12 (Engineers) two different equipments are used in the same net. The arrangement of the net-site matrix is completely arbitrary; its primary function is to depict the operational data acquired from the user.

In an operational problem, an additional list of data would be required, namely, those frequencies lying in the tuning range of the radio equipment under consideration that may not be considered for use in the frequency selection process because of national restrictions and other spectrum users. For this sample problem it will be assumed that there are no such limitations.

Now that the operational data is acquired, it is necessary to inquire about the technical data. For one thing, it is necessary to know to what degree the use of any one frequency by any one of the radio equipments will affect the use of any other frequency by any other equipment, not only at the same radio site but also at any other site in the problem.

An equipment data library was established. A series of laboratory measurements were performed on several different serial numbered transmitters of each type, and at several different transmitter frequencies. The range of measurement in the frequency domain was from 150 kcs to 1000 mcs. Throughout this frequency range, all antenna radiated narrow-band signals whose

amplitude was greater than 30 db above 1 microvolt were measured. These measurements were reduced to mathematical model form, whereby, given any transmitter fundamental frequency, the frequency of all measured transmitter spurious radiations could be calculated and a measured signal strength attached thereto.

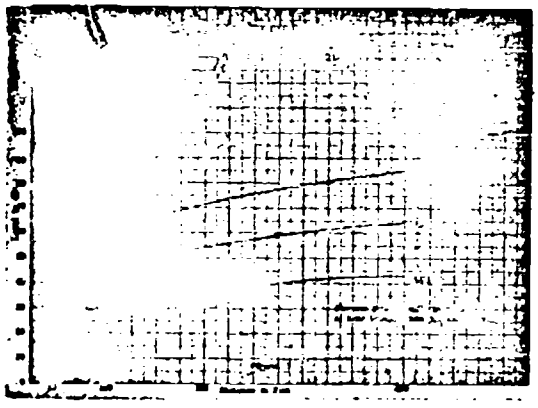
Similarly, several different serial numbered receivers of each type, tuned to several different frequencies, were subjected to a strong signal environment to determine the response of the receiver to such an environment. The frequency range of measurements was from 10 mcs to 1000 mcs. The dynamic range of signal level was from 1 microvolt to 120 db above 1 microvolt from 10 to 480 mcs and from 1 microvolt to 100 db above 1 microvolt from 480 to 1000 mcs. This signal environment simulated antenna received narrow-band signals. For each receiver response the input signal level in db above 1 microvolt of CW signal required to cause 6 db change in audio output of the "empty" receiver was recorded. This method of measuring receiver spurious responses is commonly called the "one-signal method".

The derived data was reduced to mathematical model form whereby, given the tuned frequency of any of these receivers, it was possible to calculate the frequency of all measured spurious responses. The magnitude of each spurious response was known from the measurements. Many details derived from the laboratory measurements must be neglected in this paper, since such a measurement program is a subject in itself.

After the laboratory measurement program was completed, controlled field tests were conducted to: a) correlate field results with laboratory measurements including the effects of the normal antenna systems used with these equipments; b) include the effects of nulls and reinforcements found when measuring signals near to the radiating antenna; c) correlate interference levels measured in the laboratory with physical separation of antenna systems in the field.

Altogether, more than one hundred different spurious responses were checked out in the field tests as were more than two hundred different transmitter spurious radiations and over three hundred different cases of intermodulation. Each case of interference was investigated with antenna systems separated by 50, 100, 200, and 400 feet.

For transmitter spurious radiations, it was desired to determine the probability of not hearing a spurious signal, whose level as measured in the laboratory is known, as a function of physical separation between the transmitting and receiving antennas. The average level (50 percent) and first and second standard deviations (68 and 95 percent) were of interest. Results are shown in Figure 7. Interpretation of this figure is as follows: Assume that one spurious radiation measured as described above is found to have a level of 35 db and a second spurious is found to have a level of 80 db. The first spurious (35 db) will not be heard 95 percent of the time, in a receiver located approximately 175 feet from the transmitter. If the receiver is further from the transmitter, it is even less likely that interference will occur. Similarly, the second spurious (80 db) will not be heard fifty percent of the time in a receiver located 300 feet from the transmitter. Should the receiver be placed closer to the transmitter the likelihood of interference will increase; if the receiver is further away from the transmitter the likelihood of interference will decrease.

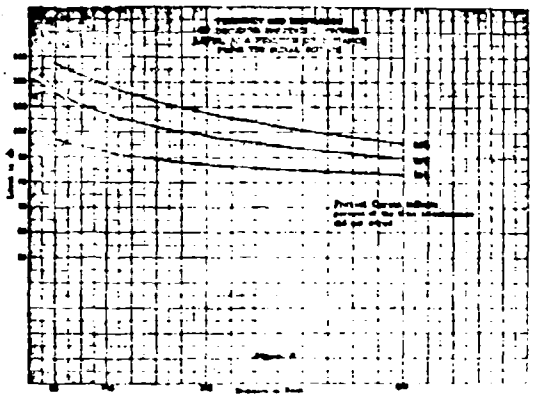


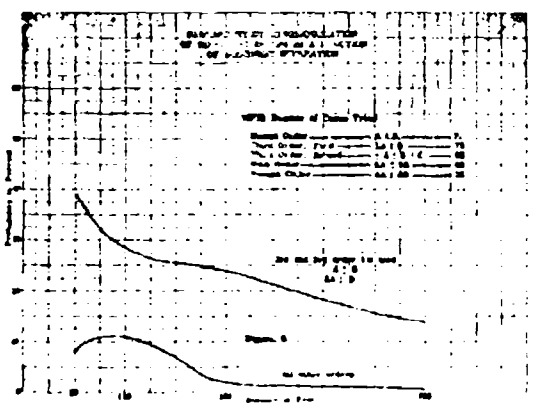
Conversely, it is possible to relate distance to the measured level of spurious radiations that must be excluded in the frequency selection process for a pre-determined reliability. For example, should it be known that average equipment separation of 200 feet is possible, exclusion of spurious radiations whose measured level was above 37 db, 54 db, or 72 db will provide 95 percent, 68 percent, or 50 percent freedom from this type of interference, respectively.

Similarly, Figure 8 indicates the results of field tests of receiver spurious responses in terms of the measured level of the responses, reliability desired, and transmitter-receiver spacings. The average system under consideration in this paper is a 150 db system i.e., 150 db path loss between the transmitter and distant receiver provides a minimum circuit. If it is assumed that 150 feet spacing between all equipments within each radio site can be achieved, all receiver spurious responses whose level was 115 db or less, and all spurious radiations whose measured level was 35 db or more, should be excluded in the frequency selection process for 95 percent surety of successful operation. The sum of transmitter spurious level (35 db) and receiver spurious response level (115 db) should be very nearly equal to the average system value (150 db) to insure balance between the two forms of interference.

Although there was a difference in transmitter power outputs, ranging from 1 watt to 100 watts, it was found that the intermodulation producing capability was the same for practical purposes regardless of the power output of the transmitters involved. It must be cautioned, however, that the presence or absence of intermodulation by order, e.g., second order, third order, etc., was of primary interest, rather than variations of amplitude within each order. Test results (Figure 9) show that second order (A ± B)

and third order (2A ± B), first kind, were the most significant. In fact, all other orders subsequently have been neglected.





Now then, for these radio equipments, it is possible to determine the inter-relationships of the various transmitters and receivers insofar as spurious radiations, responses and intermodulation are concerned. Additionally, it is possible to correlate these relationships as a function of physical spacing between the equipments.

Guard bands were also developed for each type receiver under consideration. For computational purposes, there are three guard bands. First, the guard band required between transmitter and receiver frequencies in use at the same site. This guard band is of a width determined by the R.F. selectivity of the receiver. Second, around each computed interference frequency a guard band is required to insure that the receiver will not respond to that interference frequency. The width of this guard band is a function of the I.F. selectivity of the receiver as is the third guard band. The third guard band is commonly known as the adjacent channel or co-channel guard band.

It is applicable to enable equipments operating in the same general area, but not in close proximity, to distinguish between two or more different assigned channels.

For the eight different radio equipments in this problem, thirty pages of equations were developed in the laboratory and field tests. The operational data (Figures 4, 5, and 6) and the equipment data were fed into an IBM 709 computer.

The computer was instructed to perform the following general operations:

a. Avoid all transmitter spurious radiations whose level was 35 db or greater;

b. Avoid all receiver spurious responses whose level was 115 db or stronger;

c. Avoid all second order (A ± B) and third order (2A ± B) intermodulation;

d. Generate all tentative frequencies from a random number generator;

e. Assign frequencies to the radio relay equipment before assigning frequencies to the push-to-talk nets;

f. Should, as the assignment progresses, it be found that no frequency could be assigned a particular net according to the rules, erase the partial assignment list and begin again;

g. Upon successful completion of an assignment, the list of all interference that would be created by the use of that assignment should be printed out by type of interference and by site, together with the complete assignment. Only that interference falling in the frequency ranges of the problem (1.5 - 20.0 mcs and 38.0 - 70.9 mcs) was requested.

Several hundred attempts to assign frequencies under the rules stated above were made. The computer indicated that it was highly improbable that any such list existed, since it was unable to progress further than net 14, leaving four nets unassigned.

Accordingly, rules (a), (b), and (c) above were modified to provide avoidance of spurious radiations at the 50 db level and above, spurious responses at the 100 db level and below, and to assign push-to-talk frequencies before radio relay frequencies. This scheme also proved unsuccessful in that from two to three radio relay links were not assigned frequencies.

At this point, rule (c) on intermodulation interference was modified to provide avoidance of second and third order intermodulation only when both transmitters (A and B) or one transmitter (A or B) and the receiver involved were radio relay equipments. Intermodulation generated by two push-to-talk transmitters falling into any receiver, or intermodulation generated by one push-to-talk and one radio relay transmitter falling into a push-to-talk receiver was considered "permissible" because of the intermittent nature of push-to-talk net transmissions.

Under these conditions, it was found possible to assign all frequencies except the last radio relay link (net 9). In order to determine if the sequence of frequency assignments had an appreciable effect, the computer was instructed to return to the original rule (c), that is, to assign radio relay frequencies before push-to-talk frequencies.

Complete frequency assignments were then made. A total of twenty lists for this specific problem were generated. One list is given. (Figure 10).

FREQUENCY ASSIGNMENT LIST

Net	Site	T Basic	R Basic
1	1	61.0	70.2
	2	70.2	61.0
2	2	56.1	65.9
	3	65.9	56.1
3	3	61.8	59.3
	4	59.3	61.8
4	1	64.7	62.9
	4	62.9	64.7
5	1	60.5	58.5
	5	58.5	60.5
6	2	59.9	69.2
	5	69.2	59.9
7	3	62.4	67.0
	5	67.0	62.4
8	4	54.8	68.8
	5	68.8	54.8
9	5	70.8	54.6
	6	54.0	70.8
10		40.2	
11		17.47	
12		5.66	
13		40.5	
14		17.90	
15		11.36	
16		48.1	
17		10.65	
18		18.92	

Figure 10.

A recapitulation of the limitations of these lists is in order:

a. Spurious radiations 50 db above 1 uv or greater are avoided.

b. Spurious responses 100 db above 1 uv or stronger are avoided.

c. Second and third order intermodulation wherein both transmitters or one transmitter and one receiver are radio relay equipments are avoided.

d All frequencies are generated randomly; non-randomness of the final frequency list is due to equipment limitations.

e. Radio relay frequencies are assigned before push-to-talk frequencies.

f. No other spectrum users are considered.

The resulting assignments are not interference free, but are those least likely to contain objectionable interference. In the final analysis, whether interference would exist in these assignments in operational tests or tactical situations would depend upon the density of equipments within each site. In turn, the actual density would depend upon many factors, some of which can be controlled by the user and others over which the user has no control.

It is important to recognize that this frequency selection technique evolves the best mutually compatible assignments possible with available data. Should objectionable incompatibility exists, then the real problem occurs, since it is implied that the described amount of communications is required. Therefore, if these frequency assignments are the best that are possible, but are still not sufficiently compatible, reduction in requirements, changes in doctrine, substantial improvements in equipments, or combinations of these, must be effected.

The inclusion of the requirements of other users, restrictions imposed by local and national agreements, restraints imposed by other modes of propagation, and other forms of interference will require that even stronger interference levels be "tolerated". On the other hand, better frequency management, the fact that equipment densities will vary from the statistical average, improvement in equipments and techniques - all these things will tend to alleviate the mutual interference problem.

Comprehensive testing programs, operational use of derived frequency assignments, and "time" will determine the extent to which actual field conditions can be accurately described by mathematical models.

EQUIPMENT DESIGN TRENDS IN DOPPLER-TYPE MISSILE SCORING SYSTEMS

By: B. P. Ficklin, A. H. Maciszewski, J. J. Pakan, and K. Ringer, A. R. F. Products, Inc.

Summary

Target scoring practice in missile tests has reached a considerable degree of technological sophistication, since for obvious economical reasons the destruction of the target must be avoided.

We are interested in the recording of miss-distances up to about 3000 feet and velocities between Mach 2 and Mach 10 as accurately as possible, preferably in the form of position vector with target as reference. Particular attention is concentrated on the Doppler technique because of the inherent accuracy of techniques with which frequency can be measured. Favorable test results with this technique are dependent on many equipment characteristics. The paper discusses the inaccuracies of measurements and apportions them to various effects and on that basis draws the requirements for the performance of the equipment. Recent progress made in equipment designed for this purpose is discussed.

Emphasis is placed on the design and performance of a new type of target transponder which accomplishes the required frequency translation task with adequate sensitivity (10 microvolts for 9 db $\frac{S+N}{N}$ ratio) and with sufficient rejection of spurious responses generated inside the equipment. The design includes a preselectivity filter which greatly reduces interference caused by signals of other services, not under control of the testing agency.

Certain ameliorations in the frequency layout of the entire equipment are also discussed, including methods of stabilization of oscillators which may contribute to errors in the evaluation of the Doppler shift.

1. The problem of missile scoring

This paper is concerned with a problem pertaining to the evaluation of performance of missiles with aerial targets and describes in some detail equipment used on the test ranges. The problem in question is known as "miss-distance indication" problem and it involves either the evaluation of the least distance between the missile and the target drone or the distance between them at the moment of simulated explosion. The preferable form of such information would be a position vector of the missile with the target as the origin. Usually, however, scalar distance information is considered adequate. Frequently the missile is programmed to miss the target by a predetermined distance to allow the recovery of a costly target. The least distance involved is usually less than 3000 feet while the relative velocities of the missile with respect to the target vary between about 1500 and 30000 feet per second usually not exceeding 10000 feet per second. The point of intercept may be located at about 20 to 100 miles from any ground equipment.

There are a few approaches to missile scoring. Their evaluation by Government and missile contractors continues and is complicated by the fact that no single system is judged perfect by everybody in all respects. Therefore it is probably fair to look at development of missile target scoring systems as a continuously developing field. The experience of the authors pertains to the Doppler-effect devices to which the principal part of this paper is devoted. We will, however, review briefly other principles of miss-distance measurement.

Optical tracking. It is possible to evaluate miss-distance by optical tracking using long focal length telescopes or ballistic type cameras but good result requires in general careful prior planning in the selection of the number and location of tracking stations.

Calibration by isotope radiation. Another method involves a radioisotope placed in the missile and a radiation detector placed in the target the distance being calibrated in terms of detector output. This method provides a relatively short range (approx. 150 ft) and it requires handling of radioactive materials.

Electronic devices. There is a multiplicity of electronic devices aimed at the solution of the miss-distance problem. Prior to more detailed description of the equipment which forms the basis of this paper we will mention two approaches based on different principles. One of these is the PARAMI[1] system of The Ralph Parsons Company composed of two transponders, one in the missile (different from that used for telemetry), the other in the target. The transmitting frequency of one transponder constitutes the receiving one of the other. In a system of this type a circular

oscillation is formed, the frequency of which modulates both carriers and depends in value on the delay time needed to traverse the entire loop which delay in turn depends on the propagation time of the signal and thus on the separation of the two transponders.

Another electronic system published is based on an ingenious modification of FM radar.[2] A transmitter in the target drone is modulated in frequency by a sawtooth wave and emits a signal thus modulated which after reflection from the missile reaches the receiver in the drone. A signal of low intensity from the same transmitter is also fed to the drone receiver directly and compared with the signal reflected from the missile as to the frequency. Both signals vary in frequency at a sawtooth rate; but because of the time of passage involved in the reflected signal a beat note is noticed between them at any given time. The frequency of this beat note (8 KC) is used in a closed-loop correction system to control the repetition rate of the sawtooth modulation and correct it in such a direction as to maintain the value of 8 KC of the beat note constant. Thus, for different distances between the missile and the target, different sawtooth rates will be experienced, the value of the sawtooth rate being indicative of the distance.

The above two examples show that electronic devices of quite different underlying principles can be applied to the solution of the problem of miss-distance indication.

2. The Doppler technique of miss-distance indication.

The equipment to be described is based on the principle developed by the Naval Ordnance Laboratory and used in modified form by White Sands Missile Range (Army Ordnance)[3] and is illustrated in Figure 1. The basic scheme requires a transmission source in the missile which can simultaneously be received in the target and at the ground station as shown in Figure 1. The signal from the missile received by the ground station will have a relatively constant Doppler shift while the signal received by the target will exhibit the Doppler shift the value of which will pass rapidly from positive to negative value at the time of the occurence of least distance between the two vehicles. The target transponder picks up the signal from the missile and retransmits it after suitable translation to another frequency. At the ground station the frequency of the signal retransmitted from the missile by the target is compared with that of the signal received directly from the missile. A number of frequency translations are used in the process of comparison as shown further in more detail. The difference frequency between the missile signal as retransmitted by the target and the missile signal received directly is the "Doppler shift" the value of which exhibits a rapid change of sign near the moment of least distance. The exact data of this change depend on the value of least distance and on the relative velocities of the missile and the target. The curve of the Doppler-shift versus time is obtained as an output of a FM discriminator of suitable center frequency.

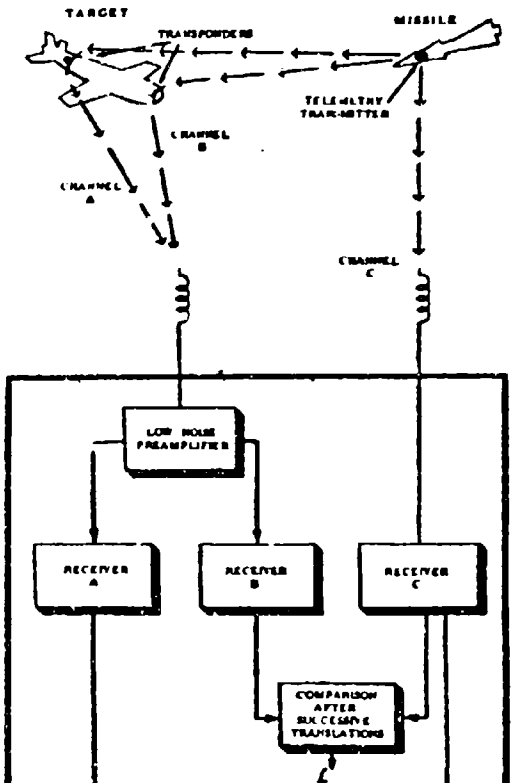
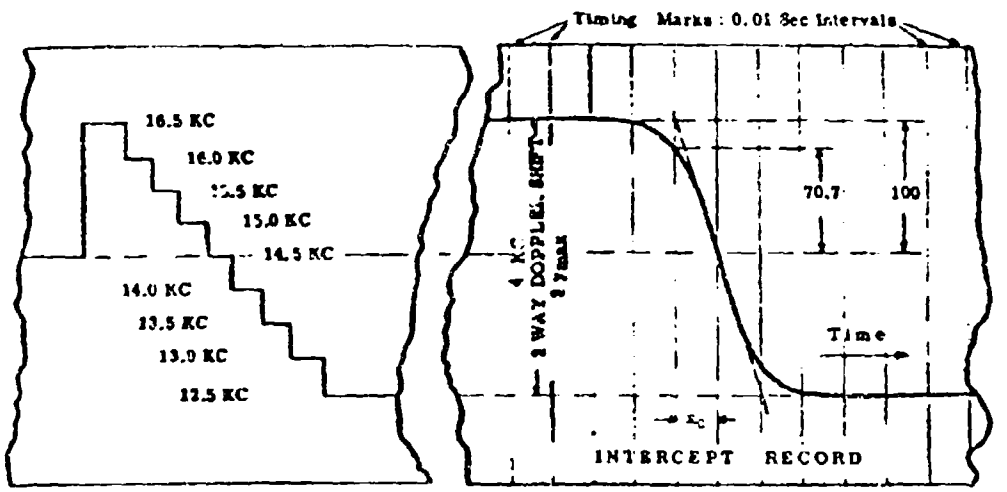


Figure 1. System Diagram of Doppler Technique of Miss-distance Indication

The output of the discriminator is calibrated in frequency by a series of crystal-controlled oscillators the frequencies of which are separated by 500 cps. The calibration frequencies are usually introduced into the discriminator immediately after the intercept. Time calibration must also be provided. Once a record of the Doppler shift for the intercept along with the calibration data is obtained (in the general form shown in Figure 2) the miss-distance may be calculated.

Essentially the Doppler technique relies on very accurate measurements of frequency and time. The necessary control of frequency and time can readily be obtained by means of standard engineering approaches to minimize the associated errors. There are many errors not usually encountered such as Doppler shifts caused by multipath signals. The two somewhat more significant sources of error are: (1) the curvature of missile path (our Doppler-shift curves are computed under the assumption that both missile and the target fly a straight course) and (2) correction needed for variations of Doppler shifts caused by the changes in the velocities of target and missile relative to the ground station near the intercept period. These errors are usually conveniently neglected which practice is justified by the fact that errors from both these sources decrease with the decreasing miss-distance. At the present time the most significant errors of the Doppler system stem from residual frequency inaccuracies of those translation oscillators the contributions of which do not cancel and from the insufficient accuracy of the data reading process because of the quality of collected data. Random frequency inaccuracies of short duration are particularly undesirable.

In this connection it should be noticed that both the Doppler-shift curve as function of time and all of its derivatives may be conveniently expressed in normalized coordinates. The normalized curves of Doppler shift and its derivative are shown in Figure 11 in Appendix I. Because of their well defined shape these curves are useful in smoothing random irregularities of obtained data. The point of least distance occurs at the point of maximum of the absolute value of the derivative curve.[7]



Example: Missile transmitter frequency = 230 MC. Total Doppler shift of 4 KC = closing velocity of approx. 8500 ft/sec. Miss-distance = 85 feet.

Figure 2. Example of Calibration Record, Intercept Record, and Miss-distance Calculation

3. Equipment approach to Doppler techniques of miss-distance indication

3.1 General data

A calculation of system performance in terms of range related to output power values of transmitters in the missile and the target and to data of sensitivity of transponder receiver and ground receiver has been made with the following results.

Assumed

Missile transmitter power (telemetry transmitter)	2 W
Target transponder sensitivity	100 microvolts in 50 ohms for 2 W transmitter output
Ground station sensitivity	2 microvolts in 50 ohms
Ground station antenna beamwidth	40°
Antenna gains in missile and in target (figure also includes estimated cable loss and loss because of differences in polarization between receiving and transmitting antenna)	- 10 db for each antenna

Results

Maximum theoretical distance (ground station to point of intercept)	300 miles
Maximum missile-drone separation for full 2 W transponder output	0.55 miles (2904 feet)

In the above calculations antenna gains are assumed with respect to a hypothetical isotropic antenna. The maximum theoretical distance between the ground station and the point of intercept is calculated assuming free space propagation without reserves for ground effect or fading. If such reserves were to be assumed the maximum distance would be about 100 miles with 90 percent reliability.[4] For line-of-sight conditions to apply the missile and the target must be at an altitude of at least 5000 feet at 100-mile distance and at an altitude of at least 50000 feet at 300-mile distance. Maximum miss-distance may increase further when the intercept point is located closer to the ground station. The sensitivity of the target transponder is very important in securing an adequate maximum value of miss-distance. For example, for the sensitivity of 1 millivolt per 2 W output the maximum separation is only 0.06 miles (316 feet) at the extreme distance to the point of intercept.

Figure 3 shows the block diagram of a typical equipment. The described system requires a transmitter in the missile but averts the necessity for adding any equipment to the missile by utilizing the telemetry transmitter already present in the missile in nearly all test firings without any modification whatsoever. In the example shown observation of Doppler shift is performed simultaneously by two transponders in the target (Channel A and Channel B). This practice is not essential and there are Doppler-MDI installations where it is not followed, the equipment in the target being limited to a single transponder. The equipment on the ground, the block diagram of which forms the major part of Figure 3, translates the Doppler shift of each channel to the center frequency of the demodulating discriminator which converts the variations of frequency into the variations of voltage. The output voltage operates a recording instrument from which we obtain the Doppler-shift curve as function of time. The scale applicable to the record of the Doppler-shift curve is obtained automatically in a manner described later.

The estimated errors of this system are tabulated in Table I. The magnitude of the Doppler shift may be quickly estimated as a function of velocity and of miss-distance from the data in Table II. The error in frequency reading, depending on the nature of its occurence in time, may effect the accuracy of miss-distance determination in various ways which may have either mutually compensating or cumulative effect. The elimination of some of these errors ordinarily occurs by the recognition of characteristic shape of the Doppler curve and by suitable smoothing of data. In order, however, to arrive at some quantitative estimate of the reasonable magnitude of allowable frequency error, let us express R_o on the basis of normalized curves of Figure 11 alternatively as:

$$R_o = \frac{v^2 f_o}{(f_D)_{max} c} = \frac{(f_{D max})^2 c}{(f_D)_{max}^2 f_o} \quad \cdots (1)$$

or as:

$$R_o = (f_{D max})\, t_o \frac{c}{f_o} \quad \cdots\cdots (2)$$

where $t_o = \frac{x_o}{k_T}$, time of flight, see Figure 2

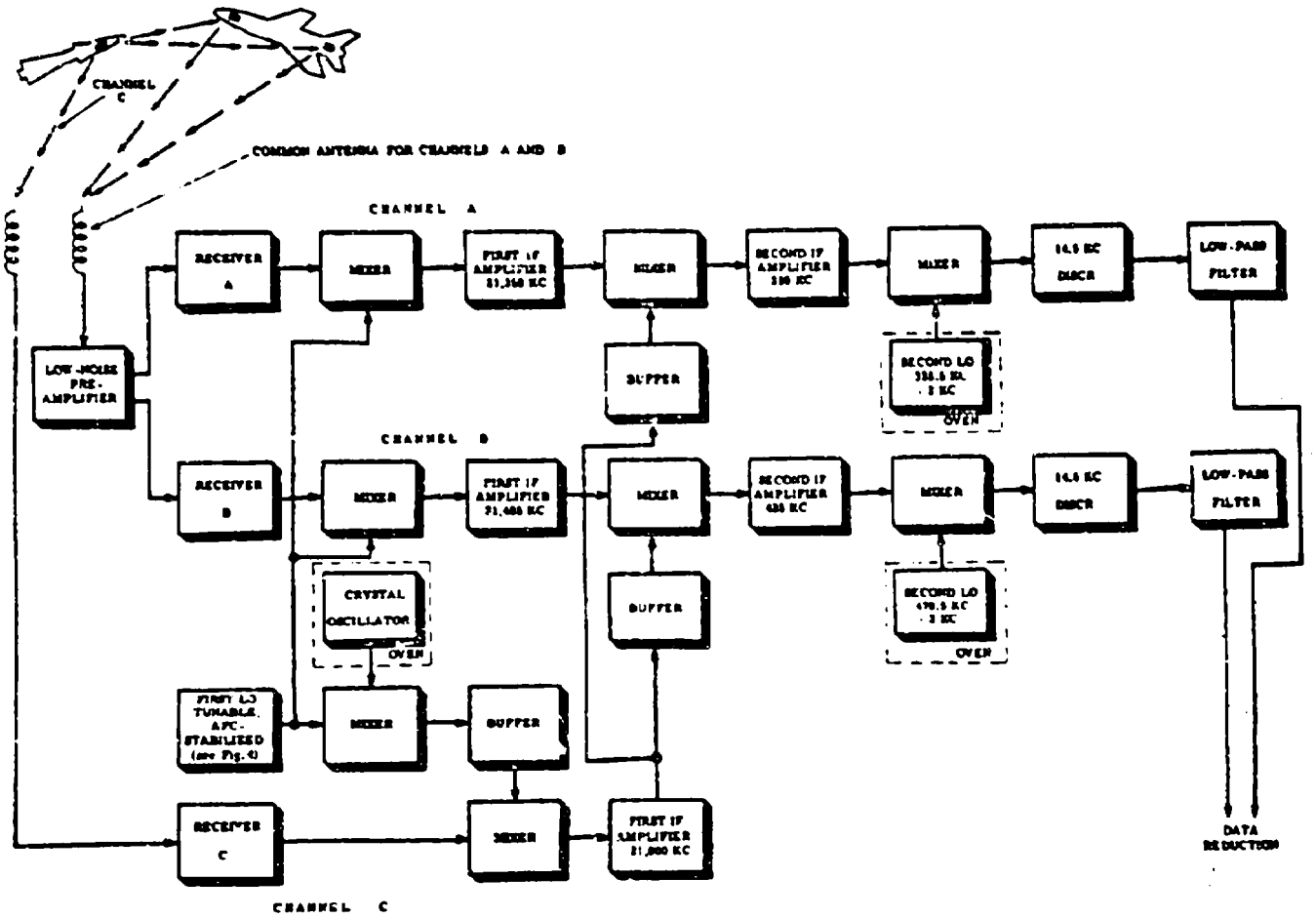


Figure 3. Block Diagram of MDI Doppler Ground Station, A. R. F. Products, Inc., Type AR-4

According to the Formulae (1) or (2) the error in $\mathcal{R}_o$ will be either percentagewise the same as error in $f_{D\,max}$ evaluation or of twice that value, depending on whether we can consider t_o or $f_{D\,max}$ as having been evaluated with accuracy. Thus, assuming the worse of the above two alternatives, we can estimate the maximum error caused by the frequency effect at relative velocity of Mach 2 as twice 16 cps referred to 905 cps (see Tables I and II) i.e. about 3.5%. This percentage error will decrease rapidly with the increase of v and of $f_{D\,max}$ becoming about 0.07 percent at relative velocity of Mach 10. In the approximation involved in (1) and (2) the error is expressed in percentages of $\mathcal{R}_o$ and does not depend on the value of $\mathcal{R}_o$.

3.2 Ground equipment

3.2.1 Receiving station

Following further the block diagram in Figure 3 we observe that the frequencies of both channels received from the target are transposed to two separate IF frequencies situated in the vicinity of 21 MC. The translation oscillator (first local oscillator) is common for both channels and stabilized by means of an AFC loop shown in more detail in Figure 4. In installations using dual target transponder more selective IF amplification is required ahead of the frequency comparison in the ground station. This greater selectivity requires a certain amount of flexibility in receiver adjustment in order to correct for frequency errors in the telemetry transmitter. Thus in these installations it was found convenient to provide for slight variation of the local oscillator. In single channel application a wider IF band pass and direct crystal control of this transposition oscillator is satisfactory.

Receivers A, B, and C were purchased as ready-made components and subsequently modified to suit the block diagram, Figure 3. A separate low-noise preamplifier was designed to enable the operation of both receivers from the common antenna. The overall noise figure of the system is about 4 db. Figure 5 shows the front panel of the ground station. It is important that the amplifiers are of sufficient bandwidth to accomodate most rapid variation of Doppler-shift curve in real time. The bandwidth needed may be calculated by conventional techniques involving Laplace transform with Doppler-shift curve approximated to a ramp function.

In view of the ambitious target for short-term frequency stability evidenced in Table I considerable attention was paid to suitable design of the constant temperature ovens in which the oscillators are contained (see Figure 3). The proportional control circuit of the oven is a bridge type and does not involve closing contacts thus preventing jumps of frequency during correction cycle. The oven is maintained at the temperature of 75 degrees C to the accuracy of 0.2 degree C.

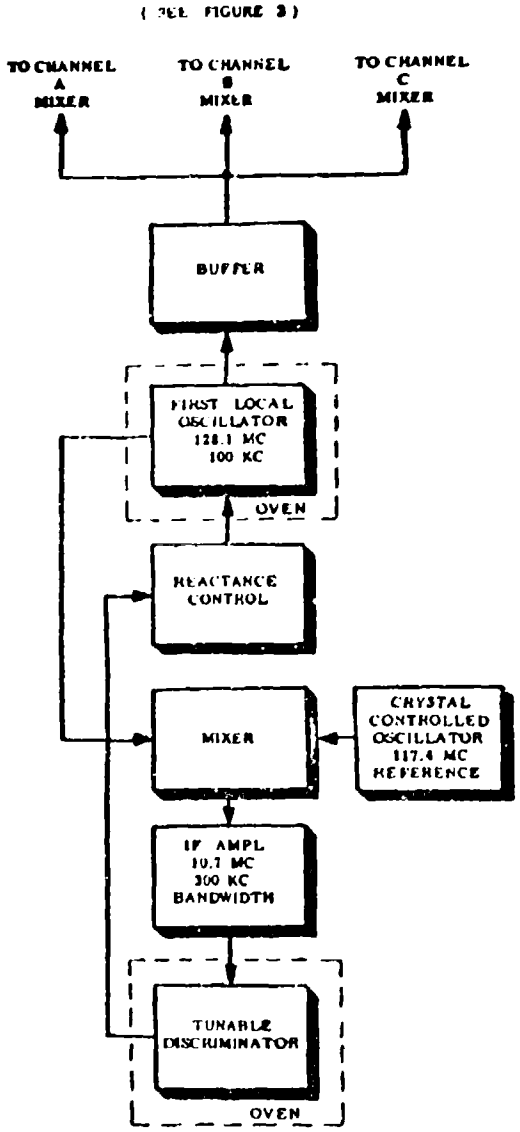


Figure 4. AFC Circuit for the First Local Oscillator

TABLE I

Frequency Error Analysis (Performance Targets)

Cause or Component	Frequency (approx) Vicinity of	Percent Error (Short term : Duration of Intercept)	Approximate Error (Cycles)
Oscillator (in the target)	80 MC	$0.185\,(10^{-8})$	15.0
Channel Transposition (on the ground, Crystal Oscillator, Fig. 3)	80 MC	$0.2\,(10^{-7})$	1.6
Last Transposition Oscillator (on the ground)	400 KC	$3\,(10^{-6})$	1.2
Calibration Crystals			1.5
Reading Error and Recording Error			5.0

Most Probable Error (Cycles):

$$\Delta F = \sqrt{(?)^2 + (1.5)^2 + (1.2)^2 + (1.6)^2 + (15.0)^2} \doteq 16 \text{ Cycles}$$

TABLE II

Observation Time, Missile Velocity, and Miss-distance Relationships

V_1 : Mach (Note 1)	2	5	10	30
$2f_{D\,max}$: CPS (Note 2) / R_o : feet (Note 3)	905	2260	4525	13575
20	18	7.0	3.6	1.2
100	90	35	18	6.0
500	450	180	90	30
1000	900	360	180	60
2000	1800	720	360	120
3000	2700	1080	540	180

Observation Time $2t_o$ (milliseconds) for $V t_o = R_o$

Observation Time $2t_o$ (milliseconds) for $V t_o = R_o$

Note 1. Missile velocity referred to sea level
Note 2. For F_o = 200 MC
Note 3. Least distance to target

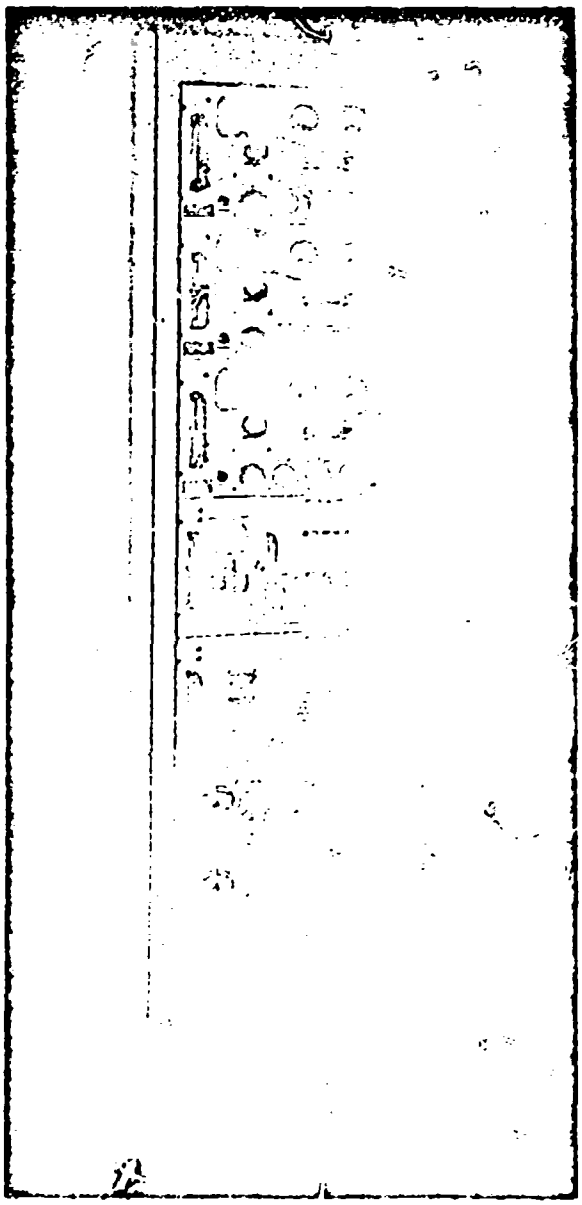


Figure 5. Ground Station, A. R. F. Type AR-4, Front Panel

Considerable attention was devoted to the problem of avoiding spurious responses by a suitable choice of circuits and numerics of frequency translations. It was found that the provision of separate intermediate frequencies was advantageous from this point of view. Frequency multiplication from crystals oscillating at low frequencies was avoided for the same reason.

3.2.2 Data reduction equipment

The present day data reduction equipment is based on the reading of the distance between the two asymptotes parallel to the abscissa axis in Figure 11 in Appendix I (reading proportional to $2 f_{D\,max}$) and on the evaluation of the slope of the plot at the point of intercept. The slope is then extended to its intersection with the asymptotes which, as it can be seen from Figure 11 in the Appendix, will occur at the points:

$$\bar{f}_D = \pm 1.0 \quad ; \quad \bar{x} = \mp 1.0$$

on the normalized curve. Once the maximum slope is correctly evaluated and the calibration constants of the recorder

k_f = ordinate/frequency

k_T = abscissa/time

are known the missile relative velocity and the miss-distance (defined as the least distance between the missile and the target) may be calculated from:

$$v = \frac{c}{F_o} \frac{y_{max}}{k_f}$$

and:

$$R_o = v \frac{x_o}{k_T}$$

where:

y_{max}, x_o } dimensions on the plot, as measured on Figure 6

k_T calibration coefficient (abscissa) in inches/sec

k_f calibration coefficient (ordinate) in inches/sec

c speed of light

F_o missile transmitter frequency

R_o least distance (miss-distance) target-to-missile

v linear velocity of the missile in the frame of reference of the target

The constants k_f and k_T are evaluated from the discriminator calibration by frequency marks and recorder speed calibration from the time marks respectively. The same basic com-

putation may be performed at better accuracy with the aid of a simple computing apparatus. This apparatus consists essentially of a film reader and a circular slide rule arrangement for the computation of the above formulae. The projector of the film reader projects simultaneously on the same screen with the plot obtained from the recorder and a set of " model "Doppler-shift curves drawn to the varying scale of abscissas. These curves vary in small steps and are labelled in terms of inverse maximum slope. The curve which fits best to the actual record is chosen for further computations performed by the circular slide rule. The vertical magnification of the projector is made variable to fit the vertical scale of the record. The above described refinement offers the advantage of curve fitting to the entire Doppler-shift plot in order to avoid errors of subjective evaluation of the maximum slope.

Figure 6 shows the actual test record obtained with the aid of the described equipment. Its data are :

f_o = 230 MC (approx)

$2y_{max}$ = 3.54 inches

$2x_o$ = 5.8 inches

k_r = 29 inches/second

k_f = 4.2 inches/KC

V = 1800 feet/second

R_o = 180 feet

The sample record is considered quite satisfactory. In particular the two extremes of the Doppler-shift curve are flat and clearly defined owing to the relatively high sensitivity of the target receiver which fact is of assistance in the evaluation of the maximum Doppler shift.

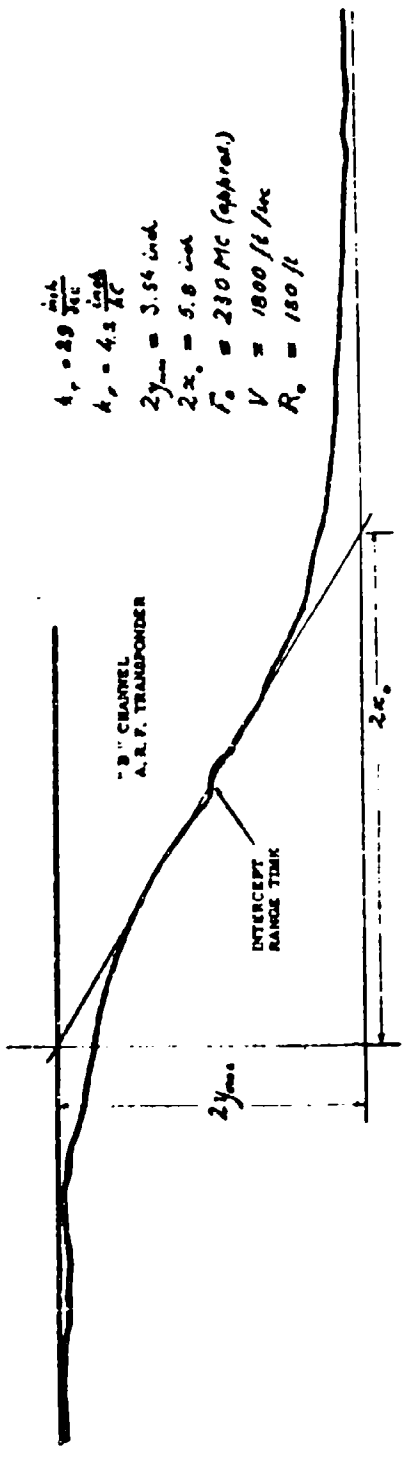


Figure 6. Actual Test Record

3.3 Airborne Equipment

Ordinary telemetry transmitter may be used in the missile. A more elaborate transceiver system is required in the drone, the block diagram of the necessary frequency transpositions is shown in Figure 7. The signal from the missile is picked-up by the receiver, amplified to a level suitable for frequency

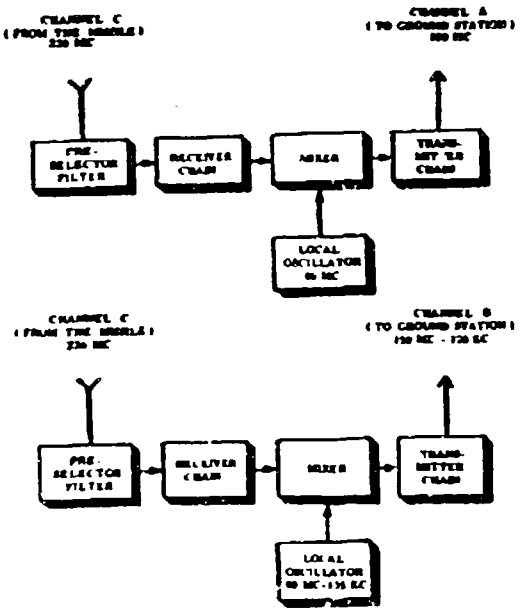


Figure 7. Block Diagram of Equipment in the Target (Two Channels Shown)

transposition, transposed to the transmitter frequency, and finally transmitted to the ground station at a power level of two watts. The following performance features of the transponder are of particular importance.

a. Frequency Stability

b. Sensitivity

c. Selectivity

d. Freedom from Spurious Responses

e. Mechanical and Environmental Features

3.3.1 Frequency stability

The frequency stability requirement is important since the contribution of the local oscillator, used for translation from receiving to transmitting frequency of the transponder, is not cancelled in the finally received record of the Doppler shift. Because of this fact we are concerned not only with good long-term stability but also with short-term variations during the intercept period of three seconds or less which could possibly be called "spurious frequency-modulation". The frequency stability (long-term) maintained is 0.0001 percent by means of a quartz crystal contained in a small oven. About twenty percent of this last figure of tolerance is contributed by this short-term instability arising from cycling of the small oven used for housing of the local oscillator.

3.3.2 Sensitivity

The sensitivity of the original apparatus used as drone equipment was approximately 1000 microvolts/50 ohms.[3] In actual flight test experience, higher sensitivity prove[d] and the sensitivity of the described equ[ipment is] 80 microvolts/50 ohms for two watts output from the transmitter. Higher sensitivity results in clearer definition of the asymptotes of the Doppler-shift curve, the distance between asymptotes being used for miss-distance determination as shown in Figure 2. Signal strength reserve is particularly needed when relative positions of missile and target-drone antennas are unfavorable.

3.3.3 Selectivity and freedom from spurious responses

The features of selectivity and absence of spurious responses should be considered jointly. From the standpoint of spurious responses generated by the equipment itself, high selectivity is needed both in the active resonant channel following the frequency transposition and in the active channel preceding it. The choice of frequencies of reception and transmission is such that, if the above condition is adequately satisfied and if sufficient shielding between various parts of the equipment is provided, there is no dangerous coincidences of intermodulation products of receiving and transmitting frequencies to be expected which would produce spurious responses of the system at a single spurious frequency.[5] The precaution of adequate shielding is, however, of great importance because of the proximity of circuit elements belonging to receiving and transmitting parts of the circuit. The most critical performance from the standpoint of rejection of spurious responses occurs for the following numerical data (approximate for illustration):

receiver channel frequency	228 MC
local oscillator frequency	78 MC
transmitter frequency	150 MC

The transmitter-chain channel should be selective enough to reject the frequency of the second harmonic of the local oscillator, i.e. 156 MC. The obtained rejection ratio exceeds 75 db.

The coaxial design of the resonators and the judicious choice of chassis shape contributed greatly to the adequacy of shielding. These contributions are difficult to express in a quantitative manner.

In order to protect the system from spurious responses arising from the intermodulation of one or more spurious signals an adequate degree of preselectivity is required. Preselectivity is defined as selectivity of circuits preceding the first active or significantly nonlinear device of the system. The reason for the provision of two tuned resonators in front of the receiver is to provide a certain amount of preselectivity. The lack of space prevents the application of an elaborate preselectivity filter but even a small improvement of this feature is of considerable value since, for example, for two spurious signals the degree of percentage intermodulation experienced is proportional to their product. Thus an equal decrease of both spurious signals percentagewise decreases the rejection ratio in the ratio of the square of its magnitude.

The basic resonant element for the described equipment was specially developed in the form of a tunable concentric resonator. The calculated Q of the concentric line portion of the resonator amounts to about 700. The measured insertion loss is approximately 3 db for the selectivity curve shown in Figure 8. According to the calculation of Fubini and Guillemin[6] this insertion loss corresponds to the unloaded Q of each resonator of about 450. The difference is probably accounted for by dielectric losses and resistance of brazed joints in the resonators. The selectivity curves of the receiving and transmitting chains are shown in Figure 8.

3.3.4 Mechanical and environmental features

Figure 9 shows the mechanical layout of the equipment. Special attention has been paid to the location of the crystal oven which is adjusted to regulate the temperature within ± 0.5 degree C. A high temperature oven (110 °C) is used in order to minimize the unavoidable difference between the temperature of the oven and the highest ambient temperature of 125 degree C. In view of very exacting temperature requirements considerable attention was paid to the problem of heat conduction away from heat sources. Figure 10 shows the main features of this design. The layout of the transponder enables easy retuning to a different channel. The transmitter output circuit which is in the form of a π filter facilitates the attenuation of harmonics of the output frequency. The equipment is designed to operate at altitudes up to 80000 feet and withstand vibrations of 10 G and shocks of 50 G.

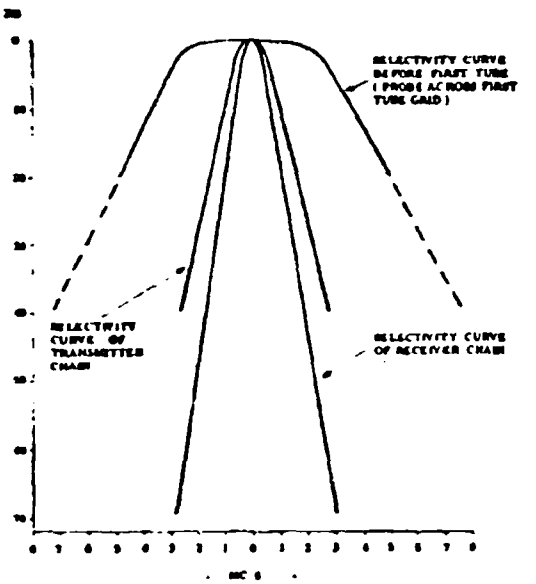


Figure 8. Selectivity Curves of Target Transponder



Figure 9. Overall View of Transponder, A.R.F. Type ART-1

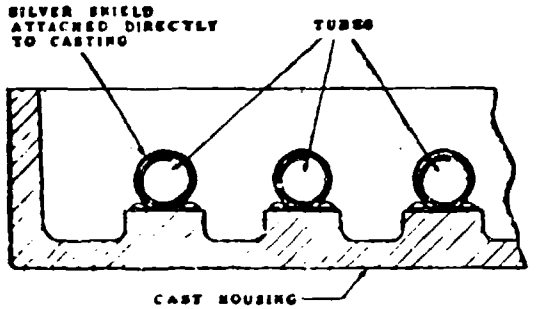


Figure 10. Transponder Housing as Heat Sink

References

1. J. A. Adams "Miss Distance Indicator Scores Missile Accuracy", Electronics, April 17, 1959 p. 43

2. W. H. Doty "Low-Cost Active Radar for Miss-distance Data", Electronics, November 20, 1959

3. P. Yaffee "Miss-distance Measuring System AN/USQ-11", Proceedings of the Fifth Annual East Coast Conference on Aeronautical and Navigational Electronics, Electronics, 1959

4. International Telephone and Telegraph Corporation "Reference Data for Radio Engineers", Fourth Edition, pp. 749-752

5. J. Hupert "Spurious Responses and Spurious Frequency Generation Arising from Frequency Translation and Multiplication", Proceedings of the Symposium on Electromagnetic Interference, 19-21 November 1957, U. S. Army Signal Research and Development Laboratory, Fort Monmouth, New Jersey

6. E. G. Fubini and E. A. Guillemin "Minimum Insertion Loss Filters", Proc. IRE, vol. 47, January 1959, pp. 37-41

7. J. Hupert "Normalized Doppler-shift Curves", Space-Aeronautics, August 1959, pp. 139, 140

Acknowledgments

The authors wish to acknowledge the help in preparation of the foregoing paper obtained from various members of the staffs of the Integrated Range Mission's Measurements Division of the White Sands Missile Range, N. M. and of the Land-Air Company's Holloman N. M. Division. In particular our thanks are due to Messrs. W. O. Patterson, A. Reynolds, and Wm. Alexander of White Sands Missile Range and to Mr. Melvin Lax of Land-Air Co. who supplied valuable counsel during design phase of our equipment and data from its field use; also to Mr. E. J. Reynolds of White Sands Missile Range who contributed information on the principles of various M. D. I. systems in use.

Acknowledgment is due to Professor J. J. Hupert of De Paul University, Chicago, research consultant to A. R. F. Products, Inc. for his advice connected with the topic of this paper.

Messrs. Jack Skolnick and Robert Schrader of A. R. F. Products, Inc. were responsible for the design of the ground station and Mr. A. Przedpelski and Mr. T. Osinski contributed to many design ideas concerning the target transponder.

The authors wish to express their appreciation to Mr. H. Richards for his painstaking attention to many details connected with the preparation of the manuscript.

Appendix I

The Doppler-shift curve and its time derivative can be conveniently presented in a normalized fashion as follows (see Figure 11).

Instantaneous frequency received is :

$$F = F_o - \frac{F_o}{c}\frac{d|\vec{r}|}{dt} \quad \text{where } c \text{ is speed of light}$$

$$|\vec{r}| = R_o \frac{1}{\cos(\arctan \frac{Vt}{R_o})}$$

If f_D = Doppler shift :

$$f_D = -F_o \frac{V}{c} \sin(\arctan \frac{Vt}{R_o})$$

The time derivative of f_D is

$$\dot{f}_D = \frac{df_D}{dt} = -\frac{V^2 F_o}{c R_o} \frac{\cos(\arctan \frac{Vt}{R_o})}{(1 + \frac{Vt}{R_o})^2}$$

Maximum value of the time derivative being

$$|\dot{f}_{D\,max}| = \frac{V^2 F_o}{c R_o}$$

If we choose the ordinates in such a manner that $t = 0$ at the moment of least distance R_o we can conveniently construct a normalized "universal" plot of f_D and $\dot{f}_D$ by using the following modified coordinates:

Abscissa (common) : $\bar{x} = \frac{V}{R_o} t$

Ordinates :

Normalized f_D : $\bar{f}_D$ = function $(\bar{x})$

where $\bar{f}_D = \frac{f_D}{\frac{V}{c} F_o}$

Normalized $\dot{f}_D$: $\dot{\bar{f}}_D$ = function $(\bar{x})$

where $\dot{\bar{f}}_D = \frac{\dot{f}_D}{\frac{V^2}{c R_o} F_o}$

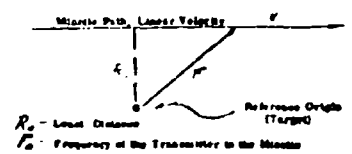
The normalized curves are shown in Figure 11. We notice that for $\bar{x} = 1$ the absolute value of the Doppler shift is 0.707 of its maximum value and

$$\frac{|\dot{f}_D|}{|\dot{f}_{D\,max}|} = \frac{1}{\sqrt{2}} \quad \text{at that point.}$$

The relationship

$$\frac{|\dot{f}_D|}{|\dot{f}_{D\,max}|} = \frac{1}{\sqrt{2}}$$

occurs at $\bar{x} = 0.5$. The curve $|\dot{\bar{f}}_D| =$ = function $(\bar{x})$ which could be generated by analogue devices owing to the existance of maximum at $t = 0$ may be useful to pinpoint the moment of miss-distance.



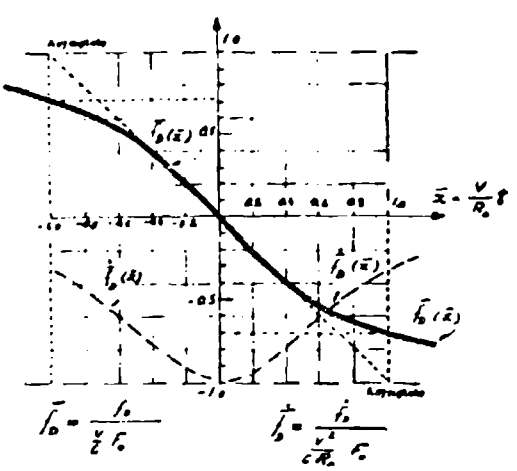


Figure 11. Normalized Doppler-shift Curves

INSTRUMENTATION SYSTEM FOR THREE-DIMENSIONAL TRACKING OF UNDERWATER MISSILES

By: C. S. Soliday & J. M. Formwalt

The development of underwater missile tracking instrumentation is reviewed briefly, and current underwater tracking ranges are described. Some discussions of the general problems and techniques of underwater tracking instrumentation is presented in relation to the current problem of instrumenting a large body of deep water for three-dimensional tracking of three or more vehicles. Some instrumentation system concepts which have been proposed to satisfy this requirement are discussed, and the basis for selection of a system for development is outlined.

A previous paper[1] surveyed the historical development of underwater missile tracking instrumentation and described the instrumentation systems in current use. The present systems that most nearly approach the desired performance are the "3-D" system developed by the Applied Physics Laboratory of the University of Washington, and the Hyperbolic Coordinate range developed by the Naval Underwater Ordnance Station.

The "3-D" system[2] incorporates one or more underwater sound projectors and four receivers mounted on a single base, the dimensions of which are small compared to the volume of water in which tracking is to be achieved (Fig. 1). A keying signal transmitted from the projector is repeated at a different frequency by a transponder installed in the missile, the repeated signal is then received by the four receivers in the array, which are located so as to define three mutually perpendicular coordinates. The round-trip transit time provides a measurement of range, while the relative phase of the signals arriving at the various receivers permits determination of the direction to the missile. This system was originally designed for use in the waters of Puget Sound at a depth of about 600 feet; a lower frequency model has since been developed and tested at 1800 foot depth, and this same model with some modifications is now being tested at still greater depth. A modification of this system is the use of synchronized clocks in the vehicle and in the instrumentation system to permit measurement of the one-way signal transit time, thereby eliminating the requirement for two-way communication.

The Hyperbolic Coordinate Range was developed in a two-dimensional form (Fig. 2) for tracking missiles in the shallow waters of Narragansett Bay. The concept involves the generation of a sound pulse in the missile which is received at three or more underwater receivers located at the corners of a large square which is intended to contain the position of the sound sources. Differences in arrival time at the various receivers determine intersecting hyperbolae to locate the position of the missile. This system does not require knowledge of the time of pulse generation for computation of the source position.

It now appears necessary to develop an instrumentation system capable of three-dimensional tracking of not only the missile but also the launching platform and the target in a large volume of deep water. The requirements for a Deep Water Range have been examined in some detail[3] and it has been proposed that the range volume (of the order of one or more miles in each of its three dimensions) be instrumented to provide overall tracking of moderate accuracy, with a small area instrumented for greater precision and a precise target-centered system for close-in relative measurements. This paper will be concerned only with the large-volume, moderate accuracy tracking system. Without specific justification, we will confine our consideration to a fixed installation with bottom-mounted instrumentation. Two general families of systems are available: a hyperbolic system, based on measurements of differences in arrival time of a signal at various points, and a spherical system, based on measurements of transit time between the missile and the receivers. Either family obviously can have two branches: an internal measurement system, where the array dimensions include the volume in which the measurements are to be made, and an external measurement system in which the array dimensions are small compared to the volume to be scanned. The APL "3-D" system mentioned above is an example of the spherical external measurement system, while the Naval Underwater Ordnance Station Rycor Range is an example of the hyperbolic internal measurement system.

Borderlon[4] of the Naval Ordnance Laboratory has compared hyperbolic and spherical systems and demonstrated the following generalizations:

1. For internal measurements, a six-receiver hyperbolic system yields the simplest and highest accuracy solution.
2. For external measurements, two cases must be considered. Close to the array, a hyperbolic system shows greater accuracy; but at distances comparable to or greater than the dimensions of the array, the spherical system is more accurate.
3. The error of the hyperbolic external system increases with the cube of the distance from the array.
4. The error of the spherical external system increases with the square of the distance from the array.
5. For either the spherical or the hyperbolic system the volume over which a specific accuracy can be attained increases as the array dimensions are increased.

Without explicit justification, we will state that the signals of interest are acoustic waves and may be either pulsed or continuous, pure tone or distributed in frequency. The principal problem is to obtain reliable detection and accurate localization throughout the volume that must be covered. We may select operating frequency or frequencies, signal power, transmission rate, and method of processing; the choices are not independent, however, since they depend upon the volume to be covered, the propagation characteristics of sea water at the range site, physical limitations of equipment, the possibility of interference with or detection by systems to

be tested, and the need for unambiguous tracking of more than one target.

We have chosen direct measurement of sound transit time using synchronized time bases (i.e., a spherical system) for our basic mode of operation. As a back up for this, a secondary mode will be provided in which time differences between arrival of the same pulse at various detectors will permit calculation of the source position without knowledge of the time of pulse transmission, using a hyperbolic system. The choice of the spherical system as the primary mode of operation minimizes the amount of data processing required, if the array is symmetrical in space. This condition of symmetry in space is difficult to achieve, however, unless we establish a rigid structure, which implies relatively short base lines and therefore measurements in a volume external to the array, and accuracy falling off with the square of the distance to the array. If we relax the requirement for special symmetry, we can use large base lines and measure inside the volume of the array. We have selected this approach, accepting the errors inherent in positioning the individual detectors. The large base line, internal measurement array will be established as accurately as is economically feasible; the locations of individual detectors will then be determined to a higher order of accuracy, and computed positions will be corrected for errors in positioning the instrumentation.

For a long base line installation, the natural choice of an array element is a vertical string of hydrophones. Since the selected site is known to have extremely low water currents, the vertical string of hydrophones can be approximated by the use of a taut-wire mooring technique in which a submerged buoy and an anchor supply the restraint. With such a system (transit time measurements from detectors in vertical strings on the boundaries of the range) any four hydrophones which do not all lie in the same straight line can uniquely define the position of a source in space.

If transit time is successfully measured, the mathematical model is essentially that of the APL "3-D" system. If for any reason we cannot measure transit time, we can employ the differences in arrival time at various pairs of detectors (i.e., hyperbolic geometry) to determine the source position, using a three-dimensional extension of the mathematical model of the HYCOR system. In either case it will be necessary to correct for departures of actual array geometry from the ideal, and for variation in sound velocity along the path of the sound ray.

The system selected employs a vertical string of hydrophones at each station, with stations located at the corners of a square. Location of the hydrophones in depth along the vertical strings depends primarily upon the sound velocity profile at location of interest. From a curve of velocity vs depth, one can compute ray diagrams from which refraction limitations can be determined.

By a graphical process that is quite simple, but sounds too involved to describe here,[3] the information derived from the ray diagrams can be used to determine the coverage of a hydrophone at any given depth for sources at any given depth. By this method a group of receiver depths have been selected such that for the assumed velocity profile any target deeper than 400 feet can be detected by at least two receivers and any target shallower than 400 feet can be detected by at least two other receivers if the range is no greater than 10,000 yards, assuming sufficient S/N ratio.

Having made the choice of a mode of measurement and an array configuration, it is necessary to decide upon a signal, which so far has been specified only as an acoustic pulse timed by a synchronous clock. The principal objective in specifying the character of this signal is to obtain reliable detection throughout the volume in which we wish to track the vehicle. To attain this objective, we may select operating frequency, signal power, transmission rate, and signal processing methods.

To permit simultaneous tracking of a number of targets by the range instrumentation but not by each other, several signal concepts were considered. Burst type transmission was ruled out because it requires storage in the vehicle of data regarding its own position unless we are to accept very infrequent data. Directional and temporal discrimination were discarded because they would require too close control of the positions and orientations of the vehicles and too much modification of fleet equipment. Continuous transmission over a broad frequency band at low power level with correlation techniques being employed in the detection system seemed more promising; but the variability of the signal in phase and amplitude after transmission through a long sea water path and the severe requirements imposed upon the instrumentation system make this approach appear unwise. The best all around approach appears to be the use of relatively narrow band transmission outside the pass-bands of the existing systems. Fortunately, it is possible to allocate four bands (to permit tracking four vehicles simultaneously) at a reasonable frequency that will provide a minimum of interference to existing equipment. It is then necessary to choose between pulsed and continuous transmission, with pulses being selected because of less precise timing and phase control requirements as well as lower average power requirement. We then calculate the signal level required for reliable detection, and are pleased to find that for the distances we have established as a design goal the power requirements are not excessive if we base our signal-to-noise ratio on sea-state noise in the selected area in a reasonably high sea state. There is another noise-limited case to be considered, however; that is the case of an interfering target between the receiver and the target of interest. Assuming the interfering target to be a destroyer traveling at high speed, we find that our tentatively established signal level is adequate at the maximum range of interest if the destroyer does not approach the hydrophone more closely than a specific distance; considering the geometry of the proposed array, we

conclude that the probability of such targets interfering with a sufficient number of receivers to cause loss of a data point is acceptably low.

In order to establish a repetition rate for our pulses, we must compromise between a desire for a high data rate and the necessity for unambiguous determination of the signal sequence. The transit time of an acoustic pulse through sea water to the maximum range of interest establishes the minimum repetition period for unambiguous tracking if there is no attempt at coding the pulses. For some vehicles the resulting data rate would be adequate, but for high-speed maneuverable vehicles it would not. We therefore select a simple pulse length code to allow us to discriminate between pulses. These pulses may all originate in a single vehicle if there is a need for a very high data rate, or they may originate in different vehicles permitting simultaneous tracking of several vehicles, each at a lower data rate.

To recapitulate, the system concept which has been synthesized is as follows: Tracking will be accomplished inside a rectangular array having vertical strings of hydrophones at its corners. Synchronized clocks will permit measurement of transit time (and therefore range) to each hydrophone, with the capability of computing position from differences in arrival times being retained as insurance against clock failure. Four relatively narrow bands of frequencies will be employed to permit simultaneous tracking of four vehicles, with pulse length coding to permit a high data rate or tracking of a greater number of vehicles at lower data rates.

Preparations are now underway for an experiment to be carried out in the selected area and designed to establish the validity of the considerations outlined here.

REFERENCES

1. "Survey of Underwater Missile Tracking Instrumentation" by D. T. Barry and J. M. Formwalt, Proceedings of IRE Vol. 47 No. 5, May 1959

2. "An Introduction to the Three-Dimensional Underwater Tracking Range" Applied Physics Laboratory, University of Washington, APL/UW/TE 58-3, 31 Jan 1958

3. "A Study of Underwater Instrumentation for a Deep Water Range" by D. T. Barry, USNUOS Consecutive No. 303, 17 Dec 1959 (Conf) (Title Uncl)

4. "Analysis of Three-Dimensional Hyperbolic Tracking Systems" by D. J. Bordelon, NOL NAVORD Report No. 2718, 10 Dec 1952

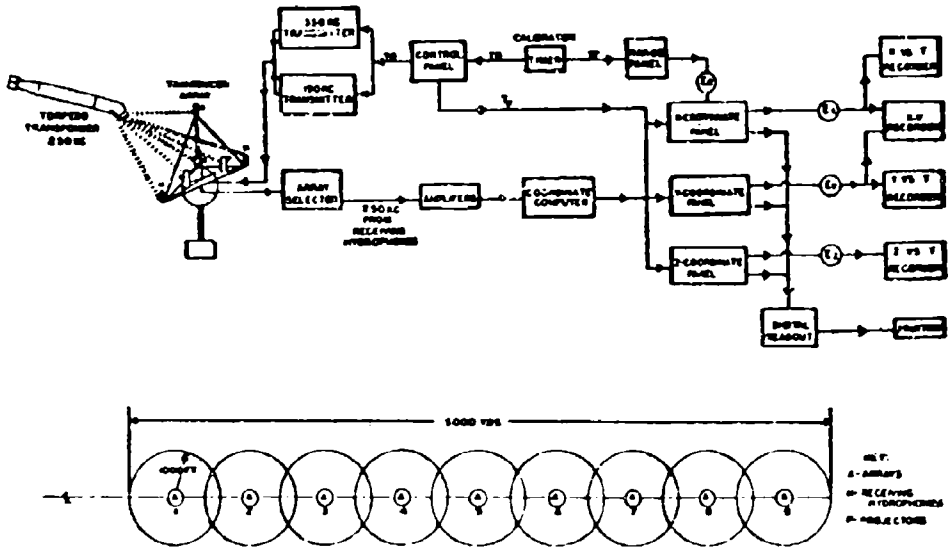


FIG #1

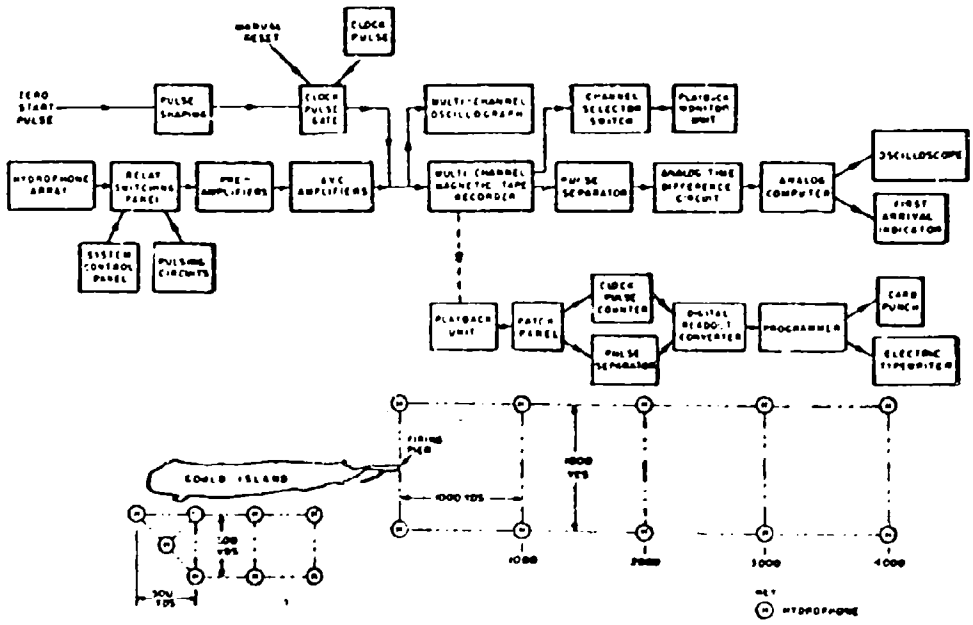


FIG #2

OPTIMUM SEARCH ROUTINES FOR AUTOMATIC FAULT LOCATION

By: S. I. Firstman, The Rand Corporation, & Brian Gluss, Armour Research Foundation of Illinois Institute of Technology

This paper comprises an extension of Model I of a recent paper of Gluss (Operations Research, July-August, 1959), and its purpose is to dictate strategies which minimize the expected cost (in time) of locating a fault in a complex system of equipment. These strategies are specialized for use with automatic testing equipment. The model assumes that the complex system consists of N modules containing n(1),... n(N) elements respectively, that the cost of examining the modules are $t_1,\ldots,t_N$ respectively, and that the costs of examining the elements within the r^{th} module are $t_{r1},\ldots,$ $t_{rn(r)}$. It is further assumed that module tests are performed to find which module is faulty before element tests are performed, and that there exist probabilities at each stage that errors of two kinds can be made: (1) that the test fails to detect an actual fault in the module or item tested; (2) that the test finds a fault that does not exist. The estimation of the probabilities of faults lying in respective modules or elements is performed in a different way from that in Gluss' paper: they are computed from element reliability data by manipulation of their λ parameters, where λ_i is the element failure rate. Furthermore, consideration is given to fault symptoms that are supplied by weighting the probabilities according to the symptom information. Because of the anticipated difficulty in obtaining the necessary parameter estimates, the analysis may be most useful for its illumination of the influence the several required estimates have on the optimum search routine.

HIGH SPEED DIRECT ELECTRONIC PRINTING CATHODE RAY TUBE

By

N. Fyler, D. Cone, R. Dorr and J. Wurts
Display Devices Department, Electron Tube Division, Litton Industries

Summary

A cathode ray tube called the PRINTAPIX* is described, which is now commercially available for direct electronic printing at high speed on non-sensitized dielectric material. The historical background, theoretical considerations, principles of operation, performance characteristics, and applications are discussed.

The high intensity, high definition electron beam used in this direct writing tube produces a charge pattern on an unsensitized dielectric surface such as paper or plastic through a unique mosaic printing head, which is an orderly array of microscopic metallic elements penetrating a thin, vacuum tight, glass face plate. The charge image, either line or continuous tone, is rendered instantaneously visible by adherence of a pigmented powder or floc, which may be permanently fixed by a rapid heat process, or erased for reuse of the base material.

Introduction

A special cathode ray tube series called Printapix has been commercially available since December 1959. Figure 1 is a photograph of representative direct writing CRT's. The purpose of this paper is to discuss some of the features of these devices.

This CRT was designed for direct, high speed, detailed, electronic printing on non-sensitized dielectric sheet such as paper. In place of the phosphor screen usually associated with direct view cathode ray tubes, a mosaic target is used as shown in Figure 2. The target consists of an orderly array of conducting elements penetrating a vacuum tight field of glass. The diameter of the conducting elements or wires is approximately 0.001 inch. Two tube types are now in production, one with 250, and the other with 500 elements per lineal inch, corresponding to 62,500 and 250,000 elements per square inch respectively.

As civilization advances, the quantity and variety of printed and pictorial material required for every day affairs increases. With the rapid progress that has been made in electronic computing and automated data-handling systems during the past decade, the need has arisen for a printing device capable of operating many times faster than electro-mechanical printers. In applications where the high speed output of information is not continuous, the demand on a printer for rapid response may be reduced through the use of intermediate, temporary storage systems, such as magnetic recording, which allow the typewriter or other electro-mechanical printer to catch up with the high speed information source while the latter is idle. Even in cases where this procedure is adequate, a direct high speed printer could reduce complexity and improve reliability.

It is not difficult to state in general terms the properties of an idealized high speed printer. The "type" must have extremely small mass, so that it can be positioned rapidly with a minimum expenditure of energy. Also, the process of making visible marks on paper or other material should consume a minimum amount of energy. These requirements suggest the use of electrical, electronic, or optical systems. A beam of light can write on photosensitive material with very little energy. However, in the formation or selection of characters with optical devices employing a mechanical system for motion, such as a small oscillating or rotating mirror, a significant amount of inertia is introduced which limits the maximum speed.

An electrical or electronic commutation system provides the ideal means for inertialess selection and positioning of characters. Fortunately, there exists a practical and efficient method of producing visible images from electric charge patterns on dielectric media.

The phenomena of electrostatic attraction and frictional, or tribo-electric charging, have been known since the time of the ancient Greeks. The concept of making a charge pattern visible by the adherence of tribo-electrically charged dust seems to be relatively old, but apparently no extensive practical use was made of it until the invention of Xerography by Carlson in 1937.[1] In the Xerographic system, an optical image is formed on a pre-charged, photoconductive plate. The resulting charge pattern is made visible by the application of dust, and the dust image is then transferred by electrostatic attraction to a sheet of paper. This system has found wide acceptance in rapid photocopying equipment. If the input data is in electrical form, it can be displayed on a cathode ray tube, imaged through a lens system to the photoconductive plate, and printed by Xerographic technique on paper.

Before proceeding to the discussion of direct electrostatic printing, it might be well to mention that there are other ways of making visible images electrically. For example, specially treated sensitized papers are used in facsimile equipment, on which elements are darkened by a flow of electrical current. The energy required to make a mark by these methods is much greater than that required to make direct electrostatic charge images.

*Trademark of Litton Industries

1. Carlson, C.F. U.S. Patent No's. 2,297,691 and 2,357,809. 1939, 1940

Direct electrostatic printing eliminates the need for the optical system, the delicate photoconductive surface, and the transfer process as used in Xerography. A needle point, held close to or touching a dielectric sheet, which is charged to several hundred volts with respect to a conductive backing, may be used to deposit a charge on the dielectric. A simple printing device, based on this principle, consists of a matrix of 35 needles arranged in a five by seven rectangular array. Each wire is switched on or off with an electronic circuit. Thus, information in digital form can be printed out as symbols made up from 35 picture elements. If more detail is required than can be reproduced with 35 picture elements, the wiring and auxiliary circuitry for a device of this kind becomes increasingly complex. The next logical development would be to bring a large number of needles through the wall of a vacuum tube, commutated with an electron beam. The Printapix printing tube, to be discussed in this paper, is a cathode ray tube with many thousands of fine wires sealed into the headplate and electronically commutated. The selection of an average elemental spacing of a few thousandths of an inch in a mosaic provides adequate resolution for the detailed reproduction of television pictures, documents, and general alpha-numeric printing.

Principles of Operation

Figure 3 shows an artist's cut-away view of a Printapix printing tube and its relation to the paper or dielectric sheet upon which images are to be printed. A high intensity, high velocity electron beam of a diameter approximating the mosaic inter-electrode spacing impinges on the minute conductive elements of the mosaic target, charging them to a negative potential. Since the conductive elements extend through the glass faceplate, this accumulation of charge can be brought in close proximity to the dielectric surface outside of the CRT. The charge pattern may be induced in the dielectric and "developed" or made visible by the adherence of a tribo-electrically charged, pigmented, plastic powder. The powder is applied uniformly in excess to the surface, and then can be shaken off or scavenged from the uncharged regions. The resulting image can be "fixed" or made permanent, by applying heat, pressure, or a transparent adhesive to bond the powder to the dielectric surface. Continuous tone reproduction is possible, because the size of the area which attracts dust is a function of the total charge on the dielectric, which, in turn, is a function of the charge on the wire over a considerable range. The reproduction of a gray scale is therefore analogous to a printed half-tone, in that the effect of gray is produced by the varying ratio of area of black to white elements. Figure 4 is a photograph of a continuous tone picture printed with a Printapix writing CRT. Figure 5 is a micro-photograph of this printing.

A two-dimensional picture can be produced in any one of several ways. Each member of the Printapix family is particularly suited for operation in one or more of the modes to be described. A tube in which the head elements are arranged in a linear array can be used in raster scan applications. The horizontal, or higher speed, sweep is an electronically generated scan along the line of wires. The paper is moved mechanically past the head of the tube to provide the equivalent of the vertical, or slower, sweep. A picture can be printed if the beam is modulated with a video signal derived from a flying spot scanner, television camera, or other pattern-generator. A non-interlaced scan is preferred. Figure 6 is a photograph of a test pattern printed in this fashion. The single line tube is also useful as an oscillograph tube, with wave-form deflection along the wires. Mechanical motion of the dielectric provides a linear time base.

A tube with an area mosaic head can be used to print two-dimensional images on a large area field, permitting random access capability. Any mode of deflection, such as rectangular scanning, polar scanning, or curve-tracing, may be used with appropriately shaped mosaics. Alpha-numeric characters can be generated by scanning, by the use of a supplementary deflection system driven by a character generator, or by the formation of a shaped beam in a special gun.

The mechanical system for use with the printing tube includes the mounting for the tube and its external deflection and focus components, the applicator or developer, the fixer, and a suitable paper transport device. The detailed design of these units may vary over wide limits, depending upon the writing speed, the type of presentation, and the disposition of the printed material as it leaves the machine. The dust may be brushed on, picked up from a cloud, or allowed to fall on the surface. The excess dust may be removed by gravitation, shaking, blowing, sucking, or brushing. The choice from among these methods is primarily a matter of convenience in a particular system. When the paper is stationary or moving slowly, fixing can be accomplished by heating the back surface of the paper to a relatively low temperature by direct contact with a heating element. In high speed systems, radiant heating of the front surface of the paper by a high temperature source will melt the plastic in the shortest possible time. Conventional methods of handling paper tape, either with continuous or discrete motion, are applicable to Printapix systems.

Electrically, these writing tubes operate in a manner very similar to other cathode ray tubes. Production tubes employing magnetic deflection are at present the most popular.

In general, deflection and focus coils which would be suitable for a particular display on a directly viewed cathode ray tube could also be used with these printing CRT's in the same type of display. Deflection circuitry can be designed to meet the needs of specific applications. A special form of this CRT is available with composite electrostatic "vernier" deflection and magnetic main deflection.

In order to keep the writing head at ground potential, it is desirable to operate the tube with the electron gun at a high negative potential. In this mode of operation, the beam current modulating signal is coupled to the grid through a high

voltage isolation device. One system of coupling, which works satisfactorily with video signals covering a frequency spectrum from dc to several megacycles, employs an rf carrier modulated by the video signal. A wide band toroidal transformer, insulated to withstand the anode voltage, is used to couple the modulator, near ground potential, to a detector and video amplifier on an insulated chassis at cathode potential. A simple capacitive coupling is sufficient in applications where good low frequency response is not required, and energy storage in the coupling capacitor can be kept small. Figure 7 is a typical circuit block diagram.

A video signal of the correct polarity to produce a positive picture on a directly viewed cathode ray tube will produce a negative picture when applied to a Printapix writing tube. Since light colored paper and dark powder are commonly used, areas of highest beam intensity will appear as black on the paper. Inverted video can be used with the printing tube to produce a positive picture. For printing symbols, line diagrams, or oscillograms, the dark lines resulting from normal positive grid drive are ideal.

Theoretical Considerations

While the general principles of electrostatic writing are readily described, a detailed explanation of the processes involved can be very complicated. The theory presented here is an attempt to explain results obtained from practical tests. The individual minute charges, small capacitances, and high values of resistivity are difficult to measure directly, particularly under dynamic conditions. The Printapix printing tube was used as a "statistical" tool. The most significant indicator of the electrical conditions was the nature of the writing obtained.

From an analysis of the writing samples, deductions can be made concerning the conditions of voltage and charge which must have existed on the head elements and the dielectric surface. Such deductions have proved useful by suggesting the direction in which certain parameters might be changed in order to improve the results, and have led to the design of a practical device.

Figure 8 is a schematic drawing of an operating Printapix CRT with an exaggerated scaling to assist in illustration. The CRT is designed to operate with 10 to 30KV anode potential relative to cathode. Because the anode is usually grounded it is appropriate to use anode potential as a reference zero potential.

The net, effective current (I_N) to a particular bombarded element consists of the beam current (I_B) minus the currents of secondary emission (I_S), target leakage (I_L), and the current to the exterior of the tube (I_P). Therefore:

$$I_N = I_B - (I_S + I_P + I_L) \qquad (1)$$

An equation for the potential of the element may be derived by commencing with the fundamental relationship:

$$V = \frac{Q}{C} \qquad (2)$$

where: V is the potential of the element; C is the total capacity of the element; and Q is the charge on the element.

The charge on the element may be expressed as follows:

$$Q = \int_{\delta}^{T_1} I_N \, dt + Q_o \qquad (3)$$

where: Q_o = initial charge condition, and

T_1 = equivalent time interval that current flows.

By substitution in equation (1) the expression for potential becomes:

$$V = \int_{t=0}^{t=T_1} \frac{I_N \, dt}{C} + \frac{Q_o}{C} \qquad (4)$$

Equation (4) indicates that the potential of the particular element is a function of its previous history and environment (Q_o), the effective capacity (C), the net effective current (I_N), and the time interval of current flow (T_1).

The current I_S may be expressed as a function of I_B:

$$I_S = \delta I_B \qquad (5)$$

where δ is the secondary emission ratio. I_L may be made very small by using a very good insulator such as glass between wire elements. Then, I_N may be expressed as:

$$I_N = I_B - \delta I_B - I_P \qquad (6)$$

In practical operation, a δ of less than unity is obtained (around 0.5).

Thus:

$$I_{N\,(typical)} = 0.5\, I_B - I_P \qquad (7)$$

Then by substitution:

$$V = \int_{t=0}^{t=T_1} \frac{0.5\, I_B \, dt}{C} - \int_{t=0}^{t=T_2} \frac{I_P \, dt}{C} + \frac{Q_o}{C} \qquad (8)$$

where T_2 is the effective time interval of I_P flow.

If the energy conveyed by the beam exceeds that lost, the element potential will proceed toward cathode potential as a limit. If energy lost from the element exceeds that gained, the potential proceeds toward ground (anode) potential.

The "rain" of a portion of the secondary electrons will tend to cause the whole target to move toward cathode potential; again, the limit

depends upon the potential energy balance.

The voltage addition of Figure 8 gives this result:

$$V = V_P + V_A \quad \text{or} \quad V_P = V - V_A \qquad (8)$$

where V_P is the effective potential near the dielectric base surface and V_A is the coupling potential drop due to contact resistance or gap discharge.

Using the same derivation technique as before:

$$V_P = \frac{Q_P}{C_P} \qquad (9)$$

and

$$V_P = \int_{t=o}^{t=T_2} \frac{I_P \, dt}{C} + \frac{Q_{P_o}}{C_P} \qquad (10)$$

where Q_P is the elemental charge on the printing surface

C_P is the elemental capacity on the printing surface

Q_{P_o} is the initial charge condition on the printing surface

T_2 is the effective time interval of current flow

In practice Q_{P_o} is zero, because an uncharged base is usually used for printing. V_A is usually very small during charge transfer, either because of direct contact conditions or because the minute gap that occurs has a very small voltage drop after the discharge starts. Thus the magnitude of V_P approaches V.

In practice, V approaches 750-2500 volts with a beam current of 100 to 200 μamps. The mosaic structure will permit a voltage gradient (per element) on the order of 250-550 volts, depending chiefly upon element size and spacing. Thus with no paper or other dielectric in position, a tube operating at full beam current will produce continuous corona.

The effective capacity of an element in the mosaic target is on the order of 0.1 to 1 μμf and the effective capacity of an elemental zone on a typical paper base is 0.01 to 0.1 μμfarads. Hence, when I_P approaches 100 μamps (I_B 200 μamps) a potential level of 500 to 1000 volts may be established on the printing base in less than 0.1 microseconds. This means that printing may be accomplished at video rates.

This potential level is adequate to tenaciously hold small particles of positively charged developing dust (pigmented resin powder) due to "coulomb" attraction, according to the well known relation

$F \text{ (Force)} = \frac{-Q_1 Q_2}{kr^2}$, because the separation (r) approaches zero when the dust is in contact.

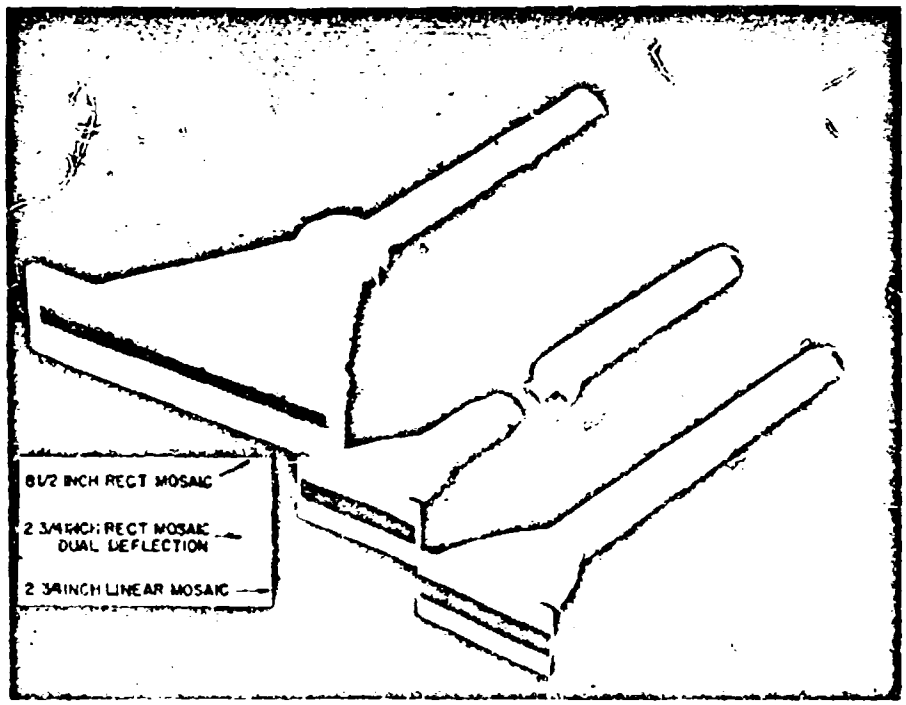
Performance Factors

The Printapix CRT is a direct efficient, high speed printer. At beam current levels that give rise to a long, useful life, printing may be accomplished at video rates. The use of 0.001 inch diameter conductive elements in close contact with the base material gives a high potential resolution level. The density of 500 elements per linear inch assures "honest" resolution levels of 150-200 "lines" per inch or over 2000 elements across an 8-1/2 inch scan. Eight shades of tone may be readily obtained.

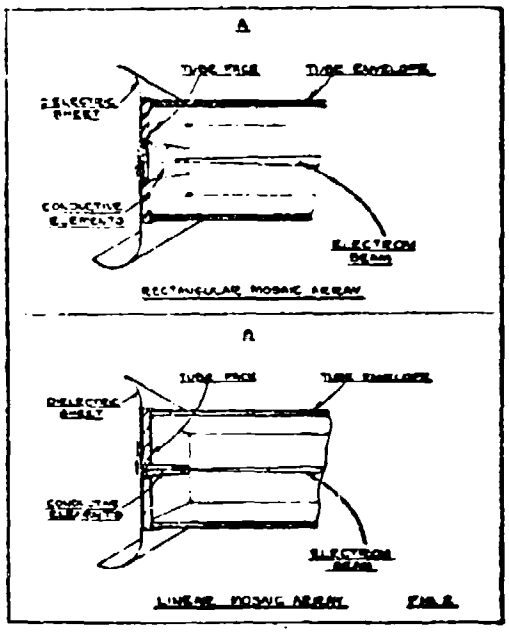
The basic principles described have been proved in practical devices. Useful advanced performance forms of Printapix writing CRT's are in development, including a very high speed alpha-numeric printer with direct digital character selection input.

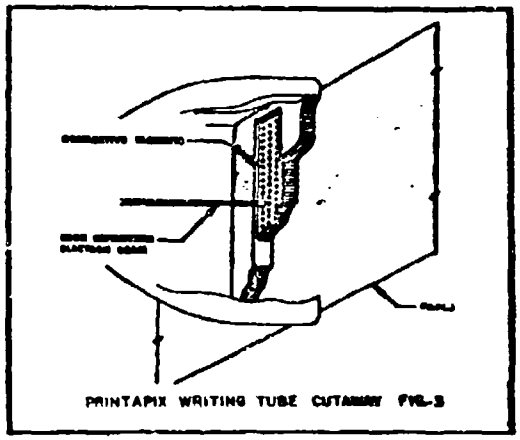
Acknowledgements

The authors wish to acknowledge the contributions of the Display Devices Department of Litton Industries. Especial mention is appropriate for Messrs. Colonel Vause, Nelson St. Jean, and William Evans.

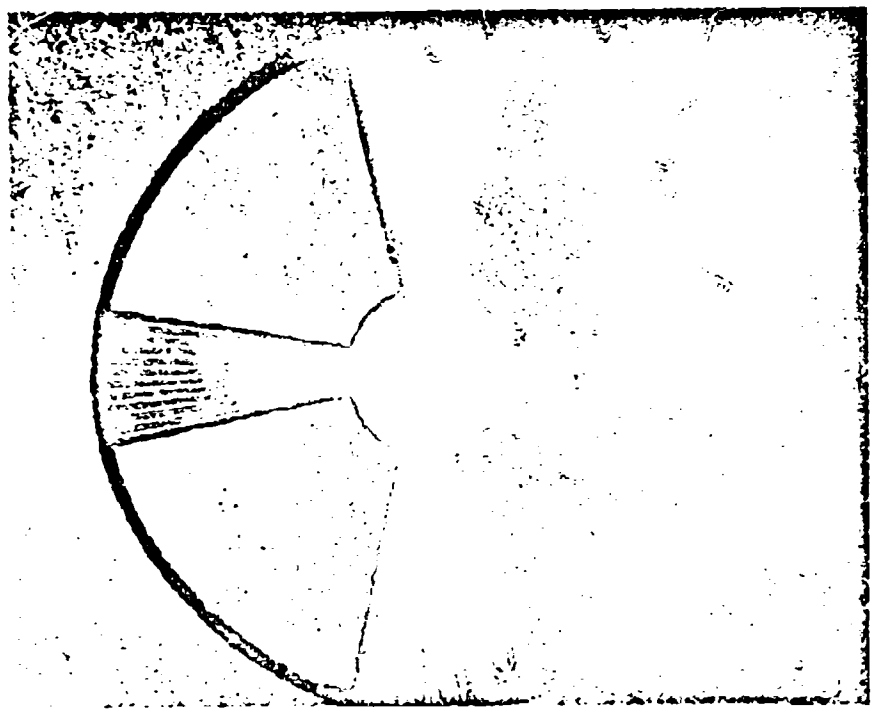


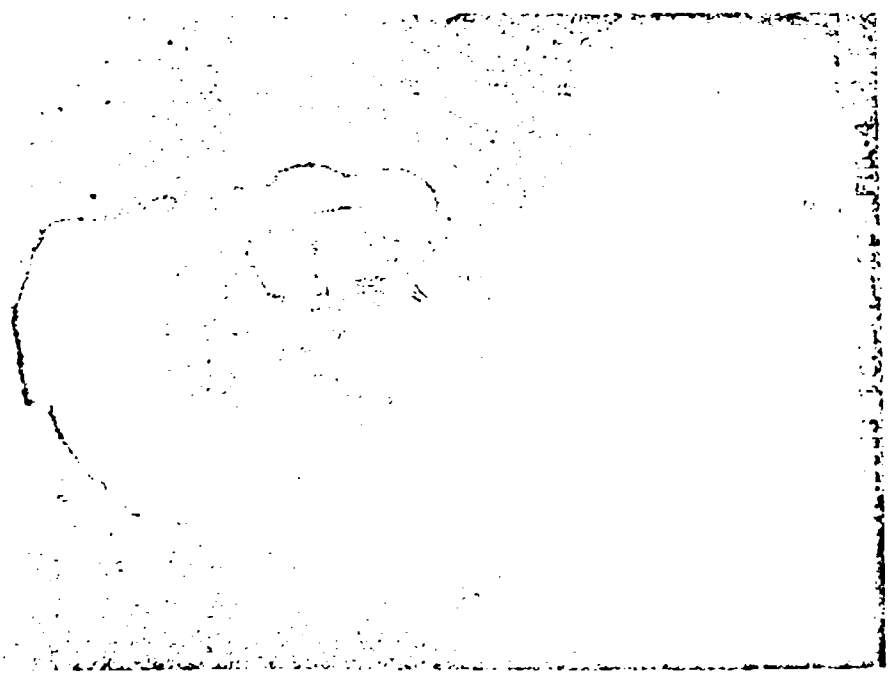
REPRESENTATIVE "PRINTAPIX" PRINTING CRT'S. FIG-1

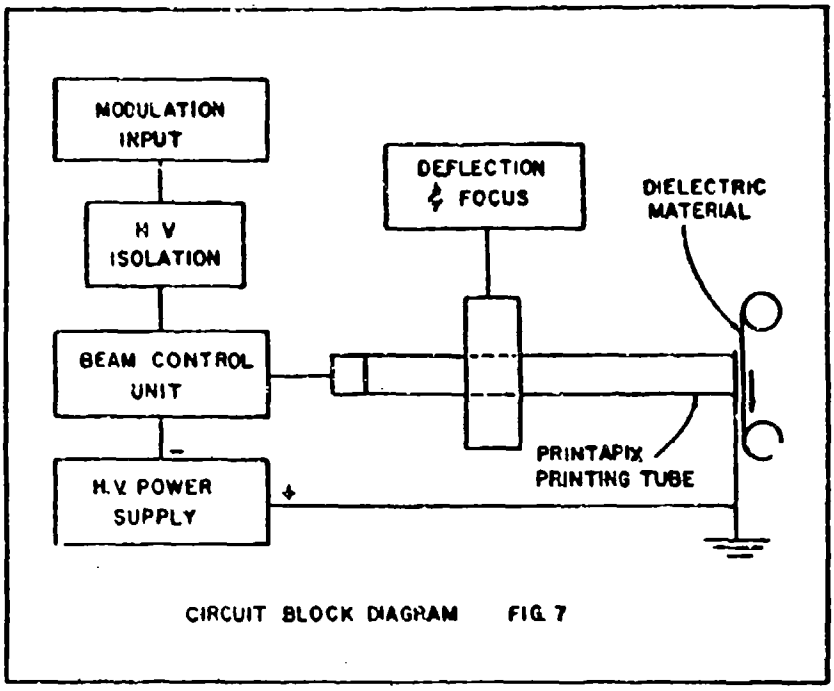




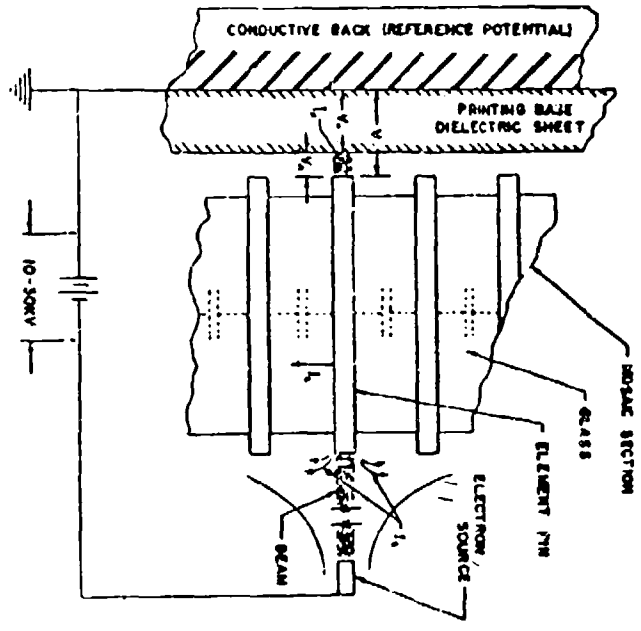
PRINTAPIX WRITING TUBE CUTAWAY FIG-3







CIRCUIT BLOCK DIAGRAM FIG. 7



SCHEMATIC OF "PRINTAPIX" CRT FIG.-8

NOTES

THE FIRMS ON THE FOLLOWING PAGES

ARE CO-SPONSORS FOR THE

1960 CONFERENCE PROCEEDINGS

OF THE

4TH NATIONAL CONVENTION ON MILITARY ELECTRONICS

NOTES

Co-Sponsors for 1960 Convention on Military Electronics

ADMIRAL CORPORATION

3800 Cortland Street — Chicago 47, Illinois

Research, Design, Development and Production of Military Electronic Systems in the fields of Communications, Countermeasures, Television, Antennas, Servomechanisms, Solid State, Radar, and ATC Systems for FAA.

A.R.F. PRODUCTS, INC.

7627 Lake Street — River Forest, Illinois

Miss-Distance Indicator Systems; Doppler tracking equipment; Voltage Comparators; Special Purpose Check Out Consoles; ECM Receivers; VHF Filters and Antenna Couplers; Test Equipment; Contract R&D.

ALFORD MANUFACTURING COMPANY

299 Atlantic Avenue — Boston 10, Mass.

RF Instruments, Antennas, Coaxial Components and Air Navigation Aids.

ASSOCIATED ENGINEERS, INC.

One Silver Street — Agawam, Massachusetts

A complete service to industry — Electronics, Research, Design, Development, Engineering, Drafting, Technical Publications plus Prototype, Production Manufacturing.

ALLIED RADIO CORPORATION

100 North Western Avenue — Chicago 80, Illinois

Everything in Electronics. Semiconductors, connectors, relays, special-purpose tubes, switches, timers, motors, resistors, capacitors, recording and sound equipment and other electronic products at factory prices.

AUTONETICS
A Division of North American Aviation, Inc.

9150 E. Imperial Hwy. — Downey, California

Product Lines: Inertial Navigation Systems, Armament and Flight Control Systems, Computers and Data Systems, Industrial Products.

Amphenol Connector Division
AMPHENOL-BORG ELECTRONICS CORP.

1830 So. 54th Avenue — Chicago 50, Illinois

Electronic Components: AN/MS, RACK & PANEL AND RF CONNECTORS: CABLE ASSEMBLIES AND HARNESSES: MISSILE AND AIRCRAFT CONNECTORS: SPECIAL CONNECTOR ENGINEERING AND PRODUCTION SERVICES.

AVCO
Research and Advanced Development

201 Lowell Street — Wilmington, Massachusetts

Research and development in: Missiles, satellites and space probes; high-temperature materials; re-entry telemetry systems; electric arc plasma generators; medical science technology.

Co-Sponsors for 1960 Convention on Military Electronics

ACF Electronics Division
ACF INDUSTRIES, INC.

11 Park Place — Paramus, New Jersey

Research-Development-Production for the Space Age; Microwave Equipment, Infrared Equipment, Navigational Display Equipment, Communications Equipment, Quick Reaction Capability, Data Processing.

BENDIX-PACIFIC
Division of Bendix Aviation Corp.

11600 Sherman Way — North Hollywood, California

Research, development, and manufacture of components and systems for Airborne Military Radar, Anti-Submarine Warfare, Electro-Mechanics, Hydraulics, Instrumentation, Missile Guidance, Systems Research, and Test Equipment.

THE BALDWIN PIANO COMPANY

1801 Gilbert Ave. — Cincinnati 2, Ohio

Manufacturer of High Accuracy Analog-to-Digital Encoders, High Frequency Preamplifiers and Telemetering Equipment, at our wholly owned subsidiary The AR & T Electronics, Inc., Little Rock, Ark.

BIRD ELECTRONIC CORPORATION

30303 Aurora Road — Solon 39, Ohio

Manufacturers of: "Termaline" Coaxial RF Wattmeters & Load Resistors, "Thruline" Directional Coaxial RF Wattmeters, "Coaxwitch" Coaxial RF Switches, and Coaxial RF Filters.

BEAVER GEAR WORKS, INC.

1025 Parmele Street — Rockford, Illinois

We are manufacturers of fine and extra fine pitch gears and small open gear trains for electronic instruments.

BUDD LEWYT ELECTRONICS, INC.

43-22 Queens Street
Long Island City 1, New York

Featuring a broad capability in R & D, and manufacturing of special purpose computers and ancillary input-output devices, displays, communications, navigation, and countermeasures.

BECKMAN INSTRUMENTS, INC.

2500 Fullerton Road — Fullerton, California

Research and products for analysis, measurement, counting and control, High speed data processing and missile checkout systems, analog computers, precision electronic components, semiconductors, medical clinical, laboratory and electronic test instruments.

BURNELL & CO., INC.

10 Pelham Parkway — Pelham Manor, N. Y.

Manufacturer of audio filters, toroidal coils, variable inductors, crystal filters, delay lines.

Co-Sponsors for 1960 Convention on Military Electronics

BURROUGHS CORPORATION

6071 Second Avenue — Detroit 32, Michigan

Computation and data processing for Missile-Guidance, Space Exploration, Airborne and Communication Systems . . . including USAF Atlas, Sage and USN AN/ASQ-8 Programs. Weapon Systems Manager for USAF Air Long Range Input Program.

CLEVITE ELECTRONIC COMPONENTS

3405 Perkins Avenue — Cleveland 14, Ohio

Designers and Manufacturers of Piezoelectric Ceramic and Crystal Transducer Elements, Ceramic Bandpass Filters, "Transfilters," self-generating Accelerometers and matching preamplifiers. Magnetic Recording Heads and Accessories.

CENTRAL ELECTRONIC, MANUFACTURERS
A Division of Nuclear Corporation of America

Power Triodes, Pulse, Rectifier, TR and Gas Noise Tubes, Ionization Gauges and Plastic Waveguides.

COLLINS RADIO COMPANY

855 35th Street, N.E. — Cedar Rapids, Iowa

Specializing in radio communication and navigation for ground, airborne and space applications. Facilities for design, production and management of complete systems.

CHANCE VOUGHT ELECTRONICS DIVISION

9314 West Jefferson — Dallas 22, Texas

Antennas
Automatic Controls
Ground Support Electronics
Navigational Electronics

CONNECTICUT TELEPHONE & ELECTRIC CORPORATION OF MERIDEN

38 Elm Street — Meriden, Connecticut

A subsidiary of the National Pneumatic Co., Inc. (and Holtzer-Cabot Division) and manufacturer and designer of electronic communication equipment and systems to military specifications.

C. P. CLARE & CO.

3101 Pratt Blvd. — Chicago 45, Illinois

Manufacturers of custom-built relays, stepping switches, and key switches for electronic and other electrical military and industrial applications.

CORNING GLASS WORKS

550 High St. — Bradford, Penna.

Manufacturers of high reliability resistors, capacitors, printed circuits (Fotoceram), trimmers, micro-circuitry, metallized glass, and ultra sonic delay lines.

Co-Sponsors for 1960 Convention on Military Electronics

COORS PORCELAIN COMPANY

600 9th Street — Golden, Colorado

The Coors Porcelain Company is depended upon for High Alumina Ceramic and Ceramic – to – Metal assemblies of prototype as well as production quantities to the Electronic Industry.

DALMO VICTOR COMPANY
Division of Textron, Inc.

1515 Industrial Way — Belmont, California

Leader in Design, Development and Manufacture of Radar Antennas of All Types, ASW Systems, Hydraulic and Pneumatic Components, Electro-Mechanical Devices, Control Systems and Servo Mechanisms.

CUBIC CORPORATION

5575 Kearny Villa Road — San Diego 11, Calif.

Missile and aircraft precision tracking systems AN/GSQ-29, (SECOR, COTAR, MOPTAR, AGAVE) range safety systems (ELSSE, Bi-COTAR), missdistance indicators (MIDAS), data handling equipment, digital instrumentation.

DAYSTROM, INC.

430 Mountain Avenue — Murray Hill, N. J.

Design, development, production; systems and subsystems, instrumentation, controller-recorders, circuit analyzers, special test, communication, fire control equipment. Servos, synchros, gyros, potentiometers and nuclear controls.

DADE TELEVISION DIVISION

West Tenth Street — Michigan City, Indiana

Complete television equipment for Military and industrial applications. Ruggedized-Transistorized airborne equipment for both closed circuit and broadcast requirements. Complete Research and Development capabilities.

DELCO RADIO DIVISION
General Motors

700 E. Firmin Street — Kokomo, Indiana

Delco Radio capabilities include research, development, manufacture of static power supplies, computers, solid state devices, communication equipment, transistorized circuits, radar, proximity fuzes, and direct conversion.

DALE PRODUCTS, INC.

Box 136 — Columbus, Nebraska

Resistors, Trimmer Potentiometers, Knobs, Hysteresis Motors and Networks. A reliable Facility for Subcontract Manufacturing of Electronic Equipment, Cables and Printed Circuits.

EDO CORPORATION
College Point, N. Y.

EDO (CANADA) LTD.
Cornwall, Ontario

World Leader in SONAR

Active – Passive – Above the Sea – On the Sea – Under the Sea

Co-Sponsors fc 1960 Convention on Military Electronics

ELECTRALAB PRINTED ELECTRONICS CORPORATION

175 "A" Street Needham Heights 94, Mass.

Printed wiring; printed circuit assemblies; exclusively for high reliability electronics. Flush circuits; CU – CON plated holes, PROTOWIRING DEPARTMENT services rush requirements. PROTOMAKA; "the do-it-yourself" laboratory.

THE GARRETT CORPORATION

9851 S. Sepulveda Blvd. Los Angeles 45, Calif.

AiResearch Manufacturing Divisions manufacturers of Systems, Packages and Components for Aircraft, Missile, Nuclear and Industrial Applications.

ELECTRONIC ASSOCIATES, INC

Long Branch, New Jersey

Manufacturers of the PACE® 231 Analog C · puter, Data Processing Equipment, Laboratory Instruments and Analog Computing Services at Princeton, N. J.; Los Angeles, California; and Brussels, Belgium.

G. B. ELECTRONICS CORPORATION
Subsidiary of General Bronze Corp.

Valley Stream, L. I., N. Y.

Design, Development and Production of Integrated Antenna Systems for Tracking, Telemetering, Communications and Space Technology.

FAIRCHILD CAMERA & INSTRUMENT CORP.

300 Robbins Lane Syosset, N. Y.

A fully-integrated Systems Management organization specializing in Aerial Reconnaissance Systems and Cameras; Rapid Film Processors; Electronic Data-gathering, Processing and Transmission Systems.

GENERAL ELECTRIC COMPANY

Industrial Park, Court Street Syracuse, N. Y.

The Heavy Military Electronics Department will feature a broad capability and competence in design and manufacture of Defense Electronics Equipment . . . from small components to the world's largest radar systems.

FENWAL ELECTRONICS, INC.

31 Mellen Street Framingham, Mass.

Pioneers in Research and Development and the Manufacture of Thermistors Precision Transducers of Temperature, Flow, Liquid Level, Thermal Conductivity Measurements; and for Temperature Compensation.

GENERAL ELECTRODYNAMICS CORPORATION

4430 Forest Lane Garland, Texas

Engaged in the development and manufacture of vidicons, storage tubes, special purpose display tubes and industrial RF equipment.

Co-Sponsors for 1960 Convention on Military Electronics

The GRAY MANUFACTURING COMPANY

16 Arbor Street — Hartford 1, Conn.

Electronic and communications products, systems and components for aviation, industry and the armed forces. Completely integrated research and production facilities to undertake wide range of projects.

THE INDIUM CORPORATION OF AMERICA

1676 Lincoln Ave. — Utica, N. Y.

Suppliers of all indium metal products – powder, foil, pellets, spheres, ingots, preforms, rod, wire, plating baths and salts.

HOOVER ELECTRONICS COMPANY

110 West Timonium Road — Timonium, Maryland

A Growing Influence in Electronics: Systems Engineering; Telemetering Components and Systems; Special Test, Automatic Checkout, Ground Support and Ground Control Equipment; Special Airborne and Mobile Instrumentation.

INSTRUMENTS FOR INDUSTRY, INC.

101 New South Road — Hicksville, L. I., New York

Designers and Manufacturers of complete Military Electronic Systems, Countermeasures, VHF/UHF Receivers, Special Receivers, Low Noise Amplifiers Wideband Systems, Telemetry and Ground Support Equipment.

HORMAN ASSOCIATES

941 Rollins Avenue — Rockville, Maryland

Electronic Instrumentation – For Government and Industry

ITT COMPONENTS DIVISION

P.O. Box 412 — Clifton, New Jersey

ITT Components Division of International Telephone & Telegraph Corporation is a major supplier of tubes, silicon and Selenium rectifiers, tantalum capacitors and hermetic seals for military and industrial requirements.

INTERNATIONAL BUSINESS MACHINES CORP. Federal Systems Division

326 Montgomery Avenue — Rockville, Maryland

Electronic air, sea, ground based information handling systems – data acquisition and application, communications, data processing. Systems management, systems development, research, engineering, production, installation and field support.

JANSKY & BAILEY, INC. (An Affiliate of Atlantic Research Corp.)

1339 Wisconsin Avenue — Washington 7, D. C.

Specialists in electronics since 1930, Jansky & Bailey, Inc. offers complete engineering services for applied research, design, and development in communications, navigation, antennas, propagation, countermeasures, and interference.

Co-Sponsors for 1960 Convention on Military Electronics

KEARFOTT DIVISION
of General Precision, Inc.

Little Falls, New Jersey

Design and manufacture of Inertial Guidance Systems, Floated Gyros, Platform Systems, Servo System Components, Ground Support and Test Equipment, Microwave Devices and Systems, Ferrite Material and Ferrite Components.

LORAL ELECTRONIC CORPORATION

825 Bronx River Avenue — New York 72, N. Y.

Designer, Developer, Producer of advanced airborne electronic systems for ASW, AEW, Navigation, Reconnaissance, Passive Detection, Penetration Aids, Bomber Defense, and Countermeasures for Department of Defense.

LEACH CORPORATION

18435 Susana Road — Compton, Calif.

Manufacturers of electro-mechanical, electronic and solid state components and systems (relays, contactors, timers & malfunction devices); instrumentation and sub-systems; power conversion and control and distribution systems.

LORD MANUFACTURING COMPANY

1635 West 12th Street — Erie, Pennsylvania

Leader in design and manufacture of vibration and shock control mountings and complete systems.

LIBRASCOPE
A Division of General Precision, Inc.

808 Western Avenue — Glendale, Calif.

COMPUTERS THAT PACE MAN'S EXPANDING MIND

LYTLE CORPORATION

1404 San Mateo Boulevard, S.E.
Albuquerque, New Mexico

Technical Publications, Slide-Soundfilms, Motion Pictures, Training Aids. Also Electronic and Electro-Mechanical Research, Development and Manufacturing including Stepping Motors, Silver-Zinc Batteries, and Printed Circuit Boards.

LITTON INDUSTRIES

336 No. Foothill Rd. — Beverly Hills, Calif.

Systems Management, Manufacture of Components, Communications Equipment, Tactical Data Systems, Electron Tubes, Inertial Guidance Systems, Geophysical Apparatus, Facsimile Equipment, Radar Systems, Computers, Instrument Landing Systems, and Space Research.

MACHLETT LABORATORIES, INC.

1063 Hope Street — Springdale, Conn.

Electron Tubes: UHF Planar Triodes; Shielded Grid Triodes, Tetrodes & Rectifiers; TV Camera Tubes; Scan Conversion Tubes; High Power Vapor Cooled Triodes.

Co-Sponsors for 1960 Convention on Military Electronics

MELPAR, INC.

3000 Arlington Blvd. Falls Church, Va.

Melpar is actively engaged in the fields of military and industrial electronics, making available electronic equipment and systems, electro-chemical materials, and printed circuit services.

NORTON ASSOCIATES, INC.

240 Old Country Road Hicksville, New York

High performance magnetic heads, record, reproduce, erase, single and multitrack for tape, film, wire, drum and magnetic ink character recognition.

MINCOM DIVISION
Minnesota Mining and Manufacturing Company

2049 S. Barrington Ave. Los Angeles 25, Calif.

Recording/Reproducing Systems. Model C-100, all-purpose instrumentation recorder, all-transistorized, instantaneous speed change. Model CV-10u, seven one-megacycle instrumentation video channels, 1/2-inch tape.

PACKARD BELL ELECTRONICS CORPORATION

12333 W. Olympic Blvd. Los Angeles 64, Calif.

Research, design, development and production of electronic equipment including automatic electronic checkout equipment, IFF, communications, radar and radio beacons, tape recorder-reproducer devices, computers, computer logic devices, ATC equipment, and high-temperature transducers and connectors.

MYCALEX CORPORATION OF AMERICA

125 Clifton Blvd. Clifton, New Jersey

Precision-molded MYCALEX glass-bonded mica, SUPRAMICA ceramoplastics and SYNTHAMICA synthetic mica. Compression-molded MYCALEX glass-bonded mica, SUPRAMICA ceramoplastics and SYNTHAMICA synthetic mica products and components.

PHILCO CORPORATION
Government and Industrial Group

4700 Wissahickon Ave. Philadelphia 44, Pa.

Research, development, design and manufacture of advanced electronics systems for space communications, radar, missile guidance and control, air traffic control and data processing, systems engineering and management.

NEMS-CLARKE COMPANY
Division of Vitro Corporation of America

919 Jesup-Blair Drive Silver Spring, Md.

Communication Receivers, Preamplifiers, Multicouplers, Display Units, Broadcast and TV Equipment, Medical Electronics and Photo Flash.

POLYTECHNIC RESEARCH & DEVELOPMENT CO., INC.

202 Tillary Street Brooklyn 1, N. Y.

PRD Electronics Co., a subsidiary of Harris-Intertype Corporation, Manufacturer of Microwave and Electronic Test Equipment.

Co-Sponsors for 1960 Convention on Military Electronics

POTTER INSTRUMENT COMPANY, INC.

Sunnyside Blvd. Plainview, L. I., New York

Digital data processing equipment: magnetic tape transports, heads, amplifiers; perforated tape readers, printers, checkout listers; system design and manufacture, including combinations of these components.

REPUBLIC AVIATION CORP.

Farmingdale, L. I., New York

Manufacturer of Advanced Aircraft, Missiles and Spacecraft.

RAYMOND ENGINEERING LABORATORY, INC.

Smith Street Middletown, Connecticut

Specialists in development and precision manufacture of Timer-Accelerometers, Programmers; Safety & Arming systems, Tape Recorders for Satellites and Missiles.

ROBERTSHAW-FULTON CONTROLS COMPANY
Aeronautical & Instrument Division

Santa Ana Freeway at Euclid Anaheim, Calif.

Manufacturer of MULTILOK communication equipment; aircraft stability augmentation amplifiers and control systems; precision crystal ovens, and industrial control instrumentation and systems.

RAYTHEON COMPANY
Microwave & Power Tube Division

Foundry Avenue Waltham, Mass.

Creative microwave technology at Spencer Laboratory in Amplitrons®, Magnetrons, Klystrons, Traveling Wave Tubes and Backward Wave Oscillators.

SPEIDEL CORP.
Industrial Division

89 Ship Street Providence, R. I.

Inertia compensated recorders for missiles and aircraft, electronic pulsers and timers, pressure switches, diaphragms and capsules, inertia switches, automatic parachute deployment devices, magneto-hydrodynamic gyro devices.

REMINGTON RAND UNIVAC DIVISION OF SPERRY RAND CORPORATION

315 Park Avenue South New York 10, N. Y.

SPERRY MICROWAVE ELECTRONIC COMPANY
Division of Sperry Rand Corporation

Clearwater, Florida

Pioneers for twenty years in Microline®, microwave test and measurement instruments, components and antennas, test sets, systems instrumentation and solid state devices.

Co-Sponsors for 1960 Convention on P[illegible]ary Electronics

SPERRY ELECTRONIC TUBE DIVISION
Sperry Rand Corporation

Gainesville, Florida

Manufacturers of klystrons and traveling wave tubes for precision radar applications – tracking, guidance, countermeasures, navigation: P to V band frequencies, milliwatt through multi-megawatt power levels.

SYSTEM DEVELOPMENT CORPORATION

2500 Colorado Ave. Santa Monica, Calif.

The System Development Corporation is an independent, non-profit organization engaged in the development of large-scale management information and command control systems.

SPRAGUE ELECTRIC COMPANY

North Adams, Massachusetts

Reliable Electronic Components and semiconductors.

TECHNICAL APPLIANCE CORPORATION

1 Taco Street Sherburne, New York

TACO Antennas for Communication and Industrial Service together with allied products will be shown in booth which will be attended by Engineering Associates.

ST. REGIS PAPER COMPANY
PANELYTE DIVISION

8218 Wisconsin Ave. Washington 14, D. C.

UNITED TRANSFORMER CORPORATION

150 Varick Street New York 13, N. Y.

- Almost 30 years experience manufacturing highest quality transformers, filters, high Q coils, magamps, etc.
- 1,000 catalog items immediately available from your local jobber
- Specials to your specifications

SUMMERS GYROSCOPE COMPANY

2500 Broadway Santa Monica, Calif.

Gyroscopic instruments and systems for manned and unmanned aircraft.

VARO MANUFACTURING CO., INC.

2201 Walnut Street Garland, Texas

RESEARCH – DEVELOPMENT – PRODUCTION in fields of:

I. MICROCIRCUITRY II. MICROWAVE
III. PRECISION POWER

Generation, Measurement, Control and Conversion.

Co-Sponsors for 1960 Convention on Military Electronics

WESTERN ELECTRIC COMPANY, INC.

195 Broadway — New York 7, New York

WHEELER LABORATORIES, INC.
A Subsidiary of Hazeltine Corp.

122 Cutter Mill Road — Great Neck, N. Y.

Consulting services, research, engineering, and design; specializing in microwave antenna development, waveguide components; and test equipment; Antenna Laboratory and test range at Smithtown, L. I.

WESTINGHOUSE ELECTRIC CORPORATION
Defense Products Group

1625 K Street, N. W. — Washington 6, D. C.

Leader in research, development, design and manufacture of military systems and equipment for the Armed Forces and the Defense Industry.

YARDNEY ELECTRIC COMPANY

40-50 Leonard Street — New York 13, N. Y.

"Pioneers in Compact Power" ®

Manufacturers of small, lightweight SILVER-CEL ® and SILCAD ® batteries.

WESTINGHOUSE ELECTRIC CORPORATION
Electronic Tube Division

P.O. Box 284 — Elmira, New York

For Information on Research and Development Activities in the Electronic Tube Field, Write to R. R. Kaemmerer, Government Contract Administrator at above address.

NOTES

AUTHOR INDEX

	Session	Page
Adelson, Leo	3.4	
Allen, George H.	4.3	308
Alsberg, D. A.	3.1	
Arm, Moses	4.1	
Armington, John C.	1.5	57
Austin, G. E.	1.3	24
Baldwin, Ruth	2.1	
Balwanz, W. W.	3.5	253
Barlow, Richard E.	2.3	111
Barta, Frank A.	3.1	
Beckhart, G. H.	2.3	117
Benanti, M.	1.3	472
Benjamin, R.	1.3	472
Berry, R. N.	3.3	214
Bickel, H. J.	1.4	488
Blom, B. V.	4.2	542
Borkan, H.	5.3	385
Brachman, Raymond J.	3.2	200
Brand, F. A.	1.3	472
Brookner, E.	1.4	488
Brouillette, Joseph W.	3.2	179
Brown, Duance C.	2.5	161
Chambers, John A.	1.2	455
Chang, Sherman Y.	2.5	521
Charton, S.	3.4	234
Chaskin, Herbert	1.4	49
Cheng, David K.	1.5	70
Clark, John W.	5.2	355
Cleeton, C. E.	3.1	
Colbert, Charles	4.5	568
Comstock, Charles C.	2.2	82
Cone, D.	5.3	623
Cope, A. D.	5.3	385
Cormier, R. J.	4.2	282
Crist, Philip W.	1.1	
Cruickshank, William J.	2.2	497
Cullen, Francis P.	4.5	331
Dardarian, Sahag	5.4	593
Dashnaw, F. J.	5.1	
Davis, James C., Jr.	3.3	212
Dean, William A.	1.5	61
Dembo, George M.	4.5	568
Dickinson, William	5.4	
Dickman, R. T.	1.1	
Dolan, George F.	2.3	120
Dorr, R.	5.3	623
Dugan, John M.	1.5	54
Easton, Roger L.	3.1	
Eichler, Ewald, Dr.	3.4	249
Finn, F. W.	2.1	
Fein, Arthur	3.4	563
Feldman, N. W.	3.4	147
Ficklin, W.	4.4	605
Firstman, Sidney I.	3.5	622
Formwalt, J. M.	4.4	618
Foster, Leroy E.	4.4	
Friedland, Irwin	3.5	258
Fyler, N.	5.3	623
Giddis, Albert R.	1.2	6
Glaser, Paul F.	5.2	361
Glass, Bruce N.	2.5	535
Gluss, Brian	3.5	622
Goodman, Louis	1.5	556
Gray, Sidney	5.3	380
Griffin, D. J.	5.5	445
Guttwein, G. K.	3.3	225
Hacking, C. A.	4.1	
Harley, T. P.	5.1	
Hawkins, Samuel R.	3.3	218
Hebbert, R. Scott	2.2	
Helf, William	4.5	331
Henry, E. F.	5.1	
Hilliard, E. J.	4.1	
Horowitz, Samuel	1.2	13
Howitt, George	5.1	
Hunter, Larry C.	2.3	111
Ikrath, Kurt, Dr.	2.2	82
Israelsen, B. P.	1.3	466
Jacobs, H.	1.3	472
Jaffe, H.	4.3	320
Jaquiss, Gerard	4.5	343
Jay, Lee A.	1.4	35
Johnson, Charles W.	3.2	179
Johnson, H. R.	1.2	451
Johnson, W.	2.1	
Jones, R. D.	5.5	420
Katz, M. D.	4.3	320
Kelleher, K. S.	1.3	30
Kennedy, P. D.	2.4	140
Keonjian, Edward	1.2	1
Klein, G.	4.1	
Kotzebue, K.	1.3	466
Kropfl, Walter	1.5	57
Kruegar, Thomas O.	3.3	203
Lanza, F. C.	2.5	517
Lambert, Louis	4.1	
Larson, Larry G.	2.5	168
Leliak, Paul	1.1	
Leoine, Michael R.	3.2	183
Leve, H. L.	4.3	302
Loeber, Carl W.	4.2	278
Lucas, R. M.	2.1	
Maciszewski, A. H.	4.4	605
Marcille, Robert E.	5.4	396

	Session	Page
Marsh, Hollock S.	4.2	282
Marx, Julius	3.2	183
Maurer, Martin	5.4	403
Maxwell, M. S.	2.1	
Mayo, Bruce R.	1.5	70
McKee, Herbert C.	1.4	49
Mendel, J.	1.3	472
Meredith, P. J.	5.5	445
Meyer, Charles H.	5.4	413
Micheli, Gene	2.5	157
Minneman, Milton J.	5.2	373
Mongello, Thomas	4.5	349
Mooney, David H.	1.4	37
Mostov, Phillip M., Dr.	5.2	575
Nelson, W. D.	2.4	153
Neuringer, Joseph L.	5.2	575
New, Ronald	2.4	133
Nordell, E. C.	2.1	
Norton, Richard L.	4.5	338
Norwood, Virginia T.	1.3	589
O'Brien, John C.	3.5	272
Otten, K. W.	5.2	367
Painter, Parker, Jr.	2.4	242
Pakan, J. J.	4.4	605
Palmer, J. E.	3.2	190
Paludan, C. T. N.	4.4	547
Parker, Sam E.	1.2	461
Parks, Douglas H.	5.5	435
Parton, James	5.3	390
Patton, R. B., Jr.	3.1	
Paulus, F. A.	5.5	445
Payne, Ronald C.	2.4	242
Perry, R. A.	1.5	54
Pinker, J.	2.1	
Pitsenbarger, G. A.	1.1	
Poland, W. B., Jr.	2.1	
Pollack, Stanley I.	4.3	294
Purl, O. T.	4.1	
Redlein, H. W.	3.1	
Redman, C. U.	3.4	230
Redwine, G. H.	5.4	413
Rennacker, H. E.	3.2	194
Richard, Victor W.	2.4	510
Rigney, Donald S.	5.2	575
Ringer, K.	4.4	605
Roberts, L. A.	1.2	451
Robinsen, Robert	1.5	57
Rosenfeld, Azriel, Dr.	3.5	265
Rosenthal, S. A.	4.3	320
Rosnagel, W. B.	4.3	315
Rothman, H. S.	2.4	503
Salonimer, D. J.	1.5	
Scharfman, W. E.	2.4	503
Schreyer, S. D.	4.1	
Schumacher, William R.	5.5	440
Schwartz, N.	3.3	214
Scully, John K.	4.5	331
Shelton, J. P.	1.3	30
Siera, Mark M.	1.4	41
Singer, J. M.	3.5	253
Skillman, William A.	1.4	37
Slagle, William C.	4.2	289
Slocum, K. W.	4.1	
Smith, Howard R.	2.2	600
Smith, J. D.	5.5	420
Smith, Robert E.	3.2	173
Soliozy, C. S.	4.4	618
St. John, G. E., Dr.	1.3	466
Stabelito, Vincent	2.5	521
Stefancek, R. T.	1.1	
Stockhoff, K. C.	2.2	75
Strossman, A. J.	2.2	75
Swanson, Nils	5.3	390
Tate, Harold N.	5.1	
Tepas, D. I.	1.5	57
Thomason, Robert L.	5.5	430
Tripp, G. P.	3.4	147
Tsou, C. H.	4.3	302
Unrah, George E.	2.3	128
Ver Wys, G.	3.4	234
Wanner, Vance R.	2.3	98
Weismann, Isaac	4.1	
Wilcox, Richard H.	2.3	98
Wingrove, E. R., Jr.	1.3	21
Winlund, E. S.	2.3	103
Wolf, J. J.	2.1	
Wong, Thomas	5.5	430
Wurtz, J.	5.3	623
Yeh, Leang P.	2.2	
Yoakum, William E.	1.5	
Zoble, Walter J.	4.4	584