UNCLASSIFIED

AD 421708

DEFENSE DOCUMENTATION CENTER

FOR

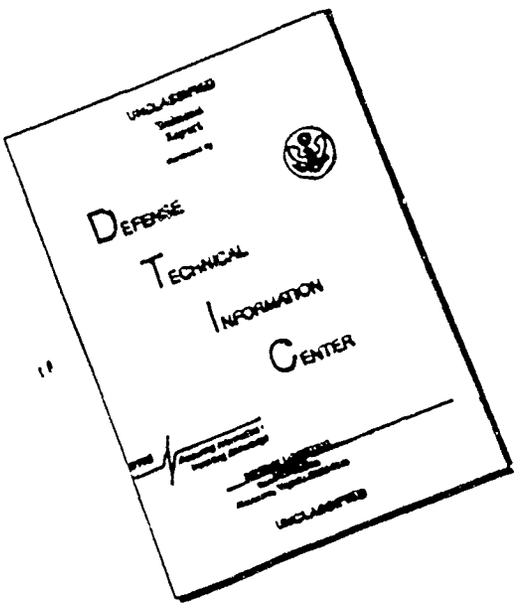
SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA



UNCLASSIFIED

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

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.

Proceedings

DDC FILE COPY 421708

ONR-16
Volume 1

338 P#5.00

The Seventh Navy Science Symposium

SOLUTION TO NAVY PROBLEMS THROUGH ADVANCED TECHNOLOGY

OCT
TISIA

Office of Naval Research
Department of the Navy
Washington, D.C.

SOLUTION TO NAVY PROBLEMS THROUGH ADVANCED TECHNOLOGY

The Seventh Navy Science Symposium

Proceedings appear in three volumes as follows:

ONR-16, Vol. 1 - Unclassified, released through Office of Technical Services, Department of Commerce

ONR-16, Vol. 2 - Classified, limited distribution

ONR-16, Vol. 3 - Classified, limited distribution

Statements and opinions contained herein are those of the authors and are not to be construed as official or reflecting the views of the Navy Department or of the naval service at large.

Copies available from OTS - $5.00

ONR-16
Volume 1

The Seventh Navy Science Symposium:

SOLUTION TO NAVY PROBLEMS THROUGH ADVANCED TECHNOLOGY

Sponsored by

THE OFFICE OF NAVAL RESEARCH

May 14, 15, 16, 1963
U.S. Naval Aviation Medical Center
Pensacola, Florida

Office of Naval Research
Department of the Navy
Washington, D.C.

ACKNOWLEDGMENTS

The success of this year's Symposium can be attributed to the efficient handling of the administrative detail and excellent management of the technical sessions. Our special thanks are extended to the following:

Dr. H. A. Imus, USNSAM
Mr. J. Y. Dowd, USNSAM
Dr. F. J. Weyl, ONR
Captain C. P. Phoebus, MC, USN, USNSAM
Dr. G. K. Hartmann, USNOL (WO)
Dr. W. P. Cooper, USNRDL
Dr. R. J. Christensen, USNEL
Mr. A. H. Schooley, USNRL
Dr. W. F. McLean, USNOTS

PROGRAM COMMITTEE

Chief of Naval Operations
Mr. T. L. Soo-Hoo

Bureau of Naval Weapons
Dr. E. S. Lamar
Dr. D. F. Bleil

Bureau of Medicine and Surgery
Captain J. P. Pollard, MC, USN
Dr. H. A. Imus

Bureau of Naval Personnel
Mr. S. Friedman

Bureau of Ships
Mr. A. W. Schmidt

Bureau of Yards and Docks
CDR E. M. Saunders

Office of Naval Research
Dr. H. Liebowitz
Captain J. F. Gustaferro
Dr. H. M. Trent
Mrs. Shirley E. Dodd
Executive Secretary
Mr. I. S. Rudin
Technical Editor
Mr. R. J. Mindak
Chairman

Host Representative
Dr. H. A. Imus
U. S. Naval Aviation Medical Center

CONTENTS
Volume I

INTRODUCTORY REMARKS vi
Dr. M. H. Schrenk, Research Coordinator, Office of Naval Research

WELCOME TO THE SEVENTH NAVY SCIENCE SYMPOSIUM vii
RADM Langdon C. Newman, MC, USN, Commanding Officer, U.S. Naval Aviation Medical Center

OPENING REMARKS ix
RADM L. D. Coates, USN, Chief of Naval Research

KEYNOTE ADDRESS xii
VADM William Schoech, Chief of Naval Material

NATO ACTIVITIES IN RESEARCH AND DEVELOPMENT xvi
Dr. I. Estermann, Chief Scientist, Office of Naval Research, London

INNOVATION: THE NEW RESOURCE xxxi
RADM Lloyd V. Berkner, USNR, Graduate Research Center of the Southwest

PRESENTATION OF THE CAPTAIN ROBERT DEXTER CONRAD AWARD xxxviii
RADM L. D. Coates, USN, Chief of Naval Research

ACCEPTANCE OF THE CONRAD AWARD xxxxi
Dr. Gerald M. Clemence, U.S. Naval Observatory, Washington, D. C.

INTERMETALLIC COMPOUND BASED MATERIALS FOR STRUCTURAL APPLICATIONS 1
W. J. Buehler, U.S. Naval Ordnance Laboratory, Silver Spring, Maryland

HUMAN QUALITY CONTROL 31
James R. Berkshire, Aviation Psychology Branch, U.S. Naval School of Aviation Medicine

PILOT PERFORMANCE DURING CENTRIFUGE SIMULATIONS OF ACCELERATION ENVIRONMENTS 41
Randall M. Chambers and W. S. Wray, CAPT, MC, USN, Aviation Medical Acceleration Laboratory, U.S. Naval Air Development Center, Johnsville, Pennsylvania

RESEARCH IN LIFE SUPPORT SYSTEMS FOR AIRBORNE AND SPACE VEHICLES ... 89
Roland A. Bosco, CAPT, MSC, USN, Air Crew Equipment Laboratory, Naval Air Engineering Center, Philadelphia, Pennsylvania

ELASTIC AND PLASTIC BEHAVIOR IN THE FINITE DEFORMATION OF METALS BY IMPACT AND EXPLOSIVE LOADING 98
Marvin E. Backman, U.S. Naval Ordnance Test Station, China Lake, California

CALOROBIC INORGANIC-ORGANIC POLYMERS 117
R. J. Landry and E. H. Bartel, U.S. Naval Ordnance Test Station, China Lake, California

PHYSICAL PROPERTIES OF THE PALLADIUM-HYDROGEN SYSTEM 129
A. I. Schindler, U.S. Naval Research Laboratory, Washington, D. C.

APPLICATIONS OF RADIOCHEMISTRY TO SOLUTION OF RADIOLOGICAL PROBLEMS 149
N. E. Ballou, U.S. Naval Radiological Defense Laboratory, San Francisco, California

CONTROL OF ELEMENT VELOCITY DISTRIBUTIONS IN SONAR PROJECTOR ARRAYS 158
D. L. Carson, G. E. Martin, G. W. Benthien, and J. S. Hickman, U.S. Navy Electronics Laboratory, San Diego, California

SOLID STATE SPECTROSCOPY OF RARE EARTH IONS 175
R. A. Buchanan, J. Murphy, and H. H. Caspers, U.S. Naval Ordnance Laboratory, Corona, California

SOLID STATE MICROWAVE AMPLIFIERS 189
Edward T. Hooper, Jr., U.S. Naval Ordnance Laboratory, Silver Spring, Maryland

PROPERTIES AND DEVICE APPLICATIONS OF EVAPORATED InSb FILMS 206
H. H. Wieder, U.S. Naval Ordnance Laboratory, Corona, California

EPITAXY OF PbS, PbTe, AND SnTe 227
Jay N. Zemel, U.S. Naval Ordnance Laboratory, White Oak, Silver Spring, Maryland

IN-PILE HALL ELECTRON TRANSPORT PROPERTY MEASUREMENTS ON ZONE-REFINED, P-TYPE SILICON 246
G. C. Bailey and C. M. Williams, U.S. Naval Research Laboratory, Washington, D. C.

SOURCES OF STRENGTH AND WEAKNESS IN GLASS REINFORCED PLASTIC STRUCTURES 263
G. R. Irwin and J. A. Kies, U.S. Naval Research Laboratory, Washington, D. C.

PROPAGATION OF HYDROMAGNETIC WAVES IN AN ATMOSPHERE OF ARBITRARY ELECTRIC CONDUCTIVITY 286
H. F. Ludloff, Pacific Missile Range, Point Mugu, California

PREFACE

We are living in an era of vanishing time. Flight speeds many times the speed of sound are becoming commonplace. Information can now be transmitted at fantastic rates through the use of high-speed communications and processing equipment. Paradoxically, more and more time is required for design and development of new, highly sophisticated hardware. We must be able to predict accurately what is possible and what might be possible. We must be in a position to control future military operations. Maximum exploitation of the discoveries and advances of our research and development program are essential in order to accomplish these tasks. This year's Science Symposium will highlight some of the activities of the Navy laboratories in the application of advanced technology to the solution of Navy problems.

It is not by chance that this Symposium is being held at one of the Navy's largest medical activities, the Naval Aviation Medical Center. Despite improvements in hardware, the basic critical sub-system is still the man, just as the basic component of any Navy laboratory is the individual scientist or engineer. The Medical Center at Pensacola is dedicated to increasing our knowledge of and ability to improve man's capabilities in the air and in outer space; we are fortunate to have the Center as our host this year. I would like to extend my personal thanks to the Program Committee, to the personnel of the Center and the Training Command, and to others who are responsible for the success of this Symposium.



L. D. COATES
Rear Admiral, USN
Chief of Naval Research

INTRODUCTORY REMARKS

Dr. M. H. Schrenk
Research Coordinator
Office of Naval Research

Good morning, distinguished guests, ladies and gentlemen. The time has come to open another Navy Science Symposium sponsored by the Office of Naval Research. I had intended to say a few words of welcome and tell you who our hosts are, but in characteristic fashion they have done it for me as you can see by the series of placards behind me on this platform. This is additional evidence of the manner in which they have been working for the past months on the many details required to insure a successful meeting. I am very appreciative of their efforts.

If you will please rise, Chaplain Kirkland will pronounce the Invocation.

I believe we have an interesting program for you, thanks to the efforts of the Program Committee. Their task was very difficult this year since of the large number of papers submitted this year only about one-third could be accepted. So if there are some among you who submitted a paper that did not find a place on the program, do not be disheartened. It was not necessarily because the Program Committee adjudged it to be inferior, but because it was impossible within the time alloted to the various sessions to accept all the good papers received.

Our Chairman for this morning's session is Dr. F. J. Weyl, Deputy Chief and Chief Scientist of ONR. Dr. Weyl is a man who has a wide knowledge of Navy science, and by virtue of a recent six-months' assignment in the Office of the Director of Defense Research and Engineering he has a fresh and an equally comprehensive knowledge of science throughout the Department of Defense.

WELCOME TO THE SEVENTH NAVY SCIENCE SYMPOSIUM

Rear Admiral Langdon C. Newman, MC, USN
Commanding Officer
U. S. Naval Aviation Medical Center

The Naval Aviation Medical Center is honored and pleased to be the host for the Seventh Navy Science Symposium, and we welcome this opportunity to get to know each of you better, and to learn from you how to do our job easier and better. At this time, I think it is only fitting that I publicly acknowledge the cooperation of the Chief of Naval Air Training, the Chief of Naval Air Basic Training and the Commanding Officer, Naval Air Station, Pensacola, whose readiness to provide facilities makes it possible for the participants in this Symposium to be accommodated.

I consider it most fitting in this day of advanced technology that the theme for this symposium is "Solutions to Navy Problems Through Advanced Technology."

This Command as well as many other Activities will be called upon to make further direct or indirect contributions to the nation's space program. The Navy has now and will have an increasing interest in all developments related to space. The exchange of techniques and the cross-fertilization of ideas make possible by gatherings of highly trained scientists such as you men are cannot help but result in an increased tempo of research developments which will be helpful to mankind in many ways. The increased tempo at which developments will be made can easily be projected by a simple comparison. It took thousands of years for man to develop the wheel. Then it took another thousand or so years for man to develop a motor to move the wheel. Next, only several decades to mass produce the automobile. Then a decade or so to produce the airplane. Now in the recent four to five years, developments have proceeded at such a rate that man is now orbiting the earth at speeds in excess of 18,000 miles per hour. This is not the end; it is probably only the beginning. As our President has said, a lunar landing will be made in this decade. Where we go from there, the future only can tell, but go we will.

I know that your primary interests are not in space technology. But whether we like it or not, space is with us and we all must play a vital part, no matter what our research interests are. Man has always dreamed of world peace. Who knows but that if all peoples became interested in space and the development of a safe means to travel in space, this could prove to be the vehicle for obtaining world peace on this planet.

The facilities that we have in the Naval Aviation Medical Center are at your disposal during your visit. We hope to make your visit most pleasant socially. There are many recreational facilities such as a twenty-seven hole golf course, swimming pools, beaches, movies, clubs and civic entertainments which you should take advantage of if the opportunity presents itself. The brochure we gave you gives many details--read it over. Let us know if we can do anything for you during your stay. It is a real pleasure to have you here.

OPENING REMARKS

Rear Admiral L. D. Coates, USN
Chief of Naval Research

I am pleased to join with the others in welcoming you to the Seventh Navy Science Symposium. Since the first one was organized in 1957, this annual series has firmly established itself as the primary means of bringing together all the Navy laboratories to discuss mutual problems and interests as well as to report on new advances accomplished by the laboratories.

This is highly important since a large segment of the Navy's efforts in research and development is carried on in its in-house laboratories. This in-house complex has been gradually built up over the past 50 years and today represents a capability that is unique in this country, if not in the world. The Naval Aviation Medical Center, our host here at Pensacola, is one example of a Navy laboratory dealing with specific problems that are not seriously attacked elsewhere. It is no exaggeration to note that the Navy's newest high-performance aircraft would have had no one to fly them if naval aviation medicine had not found means of countering the greatly increased stress involved.

The special contribution of Naval laboratories is that they are completely oriented to the sea environment and are pointedly aware of the unusual problems and requirements associated with operating in such an environment. This means that Navy civilian scientists and engineers, and the naval officers who work with them, are in a position to develop solutions to the urgent problems which are likely to confront our naval forces. The laboratories are provided with a wide variety of facilities and special equipment, most of which is not duplicated elsewhere in the nation.

None of the Navy laboratories is an individual, isolated unit, but rather each is part of the fabric of one vast research and development organization. All of the resources of this organization are freely available to the research scientist or engineer working

at any one of the laboratories. The broad exchange of scientific information, data, methods and results, such as in this symposium, is encouraged to prevent duplication of effort and to supply a cross-fertilization of ideas.

No laboratory, however, can be better than the quality of its personnel. The Navy has long recognized that a laboratory's mission cannot be accomplished unless it can obtain and hold a civilian technical staff of the highest caliber. At the same time, we are faced with the fact that this country does not have sufficient high quality scientific and technical talent to staff adequately our government laboratories, our college and university laboratories, and also the many commercial and industrial laboratories that have mushroomed in recent years. The competition is keen for top men. Yet we are not willing to lower our standards for the scientific and technical qualifications we require for our laboratory staffs.

To meet this competition, a policy was promulgated as early as 1946 that spelled out the objective of providing a working environment "comparable to that found in major academic and industrial research and development centers." We also encourage scientists at laboratories to engage in a certain amount of uncommitted or foundational research in contrast to projects or programs assigned to meet current operational needs. These laboratory-initiated projects can be in the area of either basic or applied research. This is now a line item in the Navy's RDT&E Budget.

Although much of the work in this category may have no apparent immediate application, it frequently turns out to be extremely useful at a later date. One of its main benefits is that it builds up an available store of ideas, knowlege, and techniques which can be readily applied to future problems and research tasks.

In the past few years there has been an increasing realization that Navy in-house laboratories, along with other Defense in-house laboratories, must be recognized as a major component of the military research and development effort. In the case of the Navy, these laboratories are under the technical management and direction of the technical bureaus and offices. This is characteristic of the Navy's decentralized organization designed to delegate responsibility in order to improve effectiveness.

At the same time, the Navy and other Defense laboratories should not be bogged down in administrative detail, which might hamper the urgent nature of their work. They must be effectively responsive to research and development policy worked out at the Pentagon level. These problems and the personnel problem to which I referred are of deep concern to the highest levels of the Department of Defense as well as the Navy which wants to make certain that these laboratories are properly supported and utilized.

In summary, without attempting to intrude into the domains of university and industrial research, the in-house Navy laboratory system assures us that our Navy will always be an effective force employing the most modern and effective weapons, and capable of any action which our nation may require.

It could well be said that as the Navy labs go, so goes the Navy of the future.

KEYNOTE ADDRESS

Vice Admiral William Schoech

(Read by Captain William Coleman)
Chief of Naval Material

Thank you very much.

Dr. Schrenk, Ladies and Gentlemen:

Admiral Schoech greatly regrets not being able to attend the opening of the Office of Naval Research's Seventh Annual Navy Science Symposium.

I'm sure that he has many friends in this audience, and that he has a great interest in the things that you will discuss in your meetings. Now that he is soon to move on to the post of Chief of Naval Materiel, he would have been doubly interested. But he had to testify before Congress today.

Given the choice of substituting for him here, or before Congress, I'm sure I would have chosen this.

Although I regret--as his friends here must--the necessity for a substitute, I'm glad to be here.

In this, the Seventh Navy Science Symposium conducted by the Office of Naval Research, the service research organization acknowledged by everyone to be the best, there is little that I can say to you about research as such. However, I do have some thoughts that I, as a vitally concerned consumer of your products, would like to emphasize.

These days we have a greater capability for analyzing the requirements for weapons, for assessing the utility of those weapons and for estimating the cost of those weapons, than we have ever had before. There are many methods of making these evaluations. The old tests of Suitability, Feasibility and Acceptability are a good way of stating them.

Does a proposed system do what is needed, when it is needed?

Can it be done under existing levels of technical achievement? and

Is it worth what it will cost?

My observation here is that we now need to make a more detailed analysis than ever before. Such evaluation is required both by the situation and by our superiors.

On one hand, as a nation we need a continuing technical edge in order to survive; but on the other, we can go broke by trying to purchase a technical superiority with shotgun research and development. The cost of developing weapons systems today makes it quite possible to swamp the economy we are trying to protect.

A second observation is that critical analysis of usefulness and cost should never cease. The situation is fluid. What is valid today may be wrong tomorrow.

As the Commander of the SEVENTH Fleet said recently in convening a conference of his chief lieutenants: "There is a need to beware of the status quo, to examine it, to suspect its very existence and to ignore its charm."

We must not let inertia deter regular rechecks of the requirements for what we are working on, nor let personal involvement cause us to continue a project that has been overtaken by events.

Since we are now under much closer control of the Office of the Secretary of Defense, it might be well to examine this concept in their words.

Their attitude is this, paraphrasing closely: "We categorize things as to their 'urgency' and 'feasibility.'

"There are few weapons that are so critical that without them our security would be fatally impaired. Such weapons are the nuclear bomb, the first radars, ICBM's, POLARIS, and nuclear propulsion for subs. Present critical items are an anti-missile missile and a good long range airborne submarine detection system."

(To interrupt the quote, let me expand this by saying we need an airborne capability of detecting, localizing, identifying and attacking any submarine within the horizon visible from ten thousand feet.)

To continue the quote, "The next class of urgency is 'important' (short of critical), and the third is 'marginal.'

"Cutting across these categories of 'urgency' are those of 'feasibility.' First is proven feasibility, in which existing components need only assembly to make a successful system. Second, the case where serious technical gaps need filling before a system can be completed. And third, the 'feasible' but unsuitable; that is to say unresponsive to any definite requirement supported by policy and intelligence.

"For the 'critical' and 'feasible', a crash program of concurrent development and production is indicated. For the 'important' and 'feasible', successful engineering development, followed by production, is indicated. For the 'critical' where the technology is not available, a high research rate is indicated before development and production can begin." This is the end of the paraphrasing.

On our antisubmarine problem, there can't be too great a research effort.

And that brings me to my final point. If we don't continually reassess -that is assume our responsiblities--some one will do it for us.

Or as Dr. Brown, Director of Defense Research and Engineering, put it in a speech recently:

"A group which presents, without an order of priority, a proposed increase of 40 percent for the next year of its research and development efforts, is abdicating its responsibility. Such a group is saying that <u>everything</u> it wants is overridingly important, but what <u>we</u> hear when they say that, is that <u>nothing</u> that they want is outstandingly important. Whenever a choice is not made at the <u>lower</u> level between two projects aimed at doing the same thing, between which a comparison can be made which shows that one or the other is better, the choice will be made at a higher level." He continues: "Responsibility is identical with the exercise of choice judgment. Failure to exercise such choice and such judgment is the <u>abdication</u> of responsibility, and when the <u>responsibility</u> is kicked upstairs, the <u>authority</u> goes with it." Those are strong words and he is in a position to make them stick.

R&D is <u>growing</u>, both as a percentage of the federal budget and as a percentage of the GNP. It is also growing just about as fast as there are technically qualified personnel coming along to do it. This means that we must live within our resources--and this means exercising the self-discipline that will retain for us the prerogative to decide.

Knowing you gentlemen as I do, I look to the future with great confidence. Best of luck in your efforts here at your Symposium.

NATO ACTIVITIES IN RESEARCH AND DEVELOPMENT

Dr. I. Estermann

Chief Scientist

ONR London

The North Atlantic Treaty is not only a defensive military alliance but has also the aim of developing political economic, social, cultural and scientific cooperation between its members in order to strengthen the Western world in the face of the ever-existing threat from the East. After the appearance of the first Sputnik, which demonstrated clearly that Russian science and technology could no longer be considered as a negligible quantity, The North Atlantic Council* decided to give more formal recognition to science as a major contributor to military and economic strength by the organization of a Science Committee as one of the major units of NATO. At the same time, an outstanding scientist was appointed as Science Adviser to the Secretary General and as chairman of the Science Committee. He was later elevated to the rank of Assistant Secretary General for Scientific Affairs and his office became the Division of Scientific Affairs.

As in the case of many other large organizations, especially those which have grown over a long period of time, the organization chart is a compromise between logic, history, and expediency. Thus not all the scientific activities of NATO come under the cognizance of the Division of Scientific Affairs. In keeping with the predominantly military character of NATO, the Military Committee and its Standing Group have a direct interest in some of the scientific activities. In addition, the Division of Production, Logistics and Infra-structure has certain responsibilities for research and development. The most important relationships between the scientific organizations, the North Atlantic Council and the NATO Secretariat are shown in Fig. 1.

The major components of NATO dealing with research and development will be described in broad terms on the following pages.

* The North Atlantic Council, composed of representatives of all 15 NATO nations, is the governing body of NATO

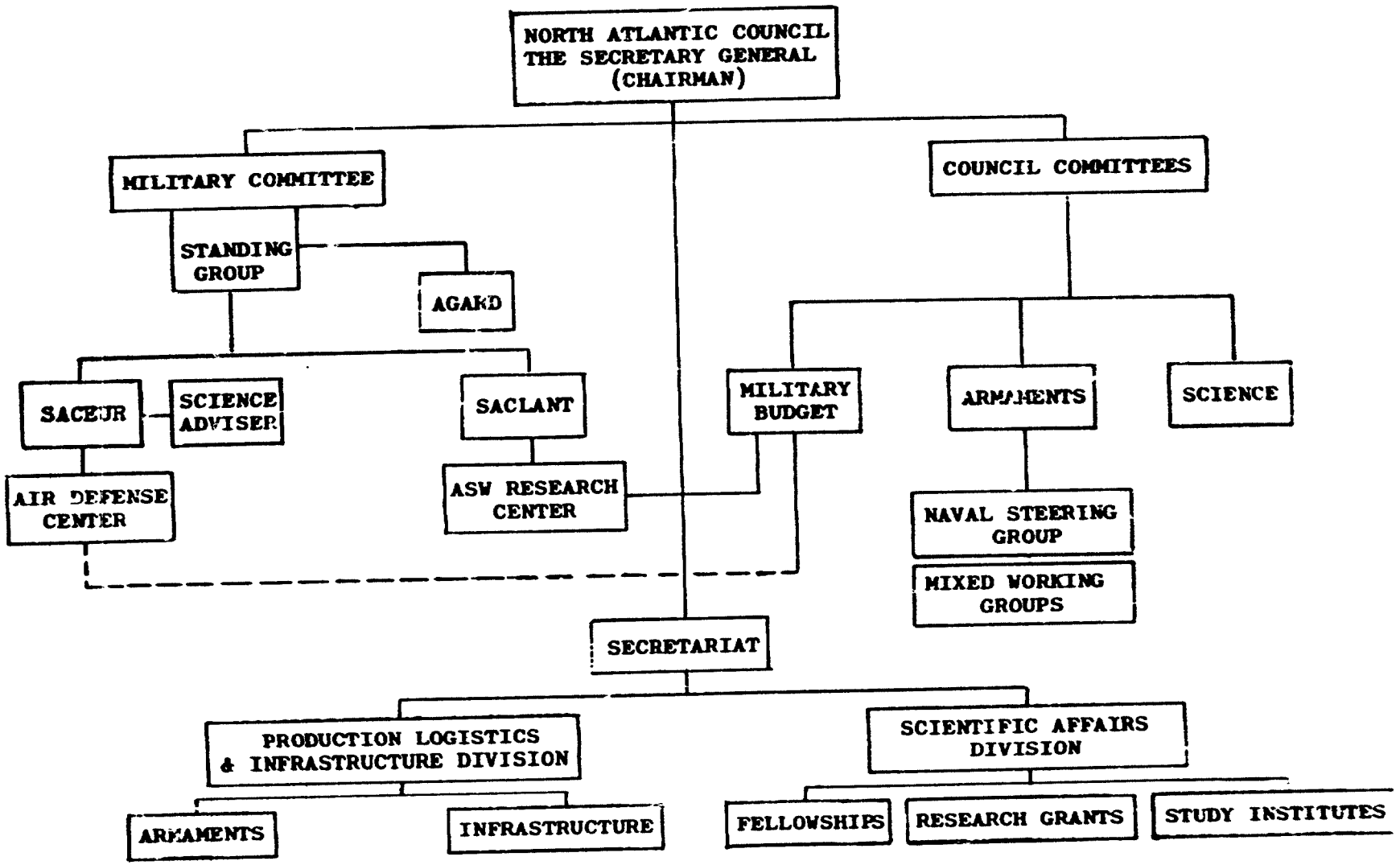

NORTH ATLANTIC COUNCIL
THE SECRETARY GENERAL
(CHAIRMAN)
MILITARY COMMITTEE
STANDING GROUP
AGARD
SACEUR
SCIENCE ADVISER
SACLANT
AIR DEFENSE CENTER
ASW RESEARCH CENTER
COUNCIL COMMITTEES
MILITARY BUDGET
ARMAMENTS
SCIENCE
NAVAL STEERING GROUP
MIXED WORKING GROUPS
SECRETARIAT
PRODUCTION LOGISTICS & INFRASTRUCTURE DIVISION
SCIENTIFIC AFFAIRS DIVISION
ARMAMENTS
INFRASTRUCTURE
FELLOWSHIPS
RESEARCH GRANTS
STUDY INSTITUTES

Figure 1

THE SCIENCE COMMITTEE AND THE DIVISION OF SCIENTIFIC AFFAIRS

The Division of Scientific Affairs is under the direction of the Assistant Secretary General for Scientific Affairs who also serves as the chairman of the Science Committee. The Science Committee, which meets three or four times per year, is composed of representatives from the 15 member nations, all of them prominent scientists. It provides broad policy direction, discusses the budget of the Division for Scientific Affairs, and formulates programs aimed at strengthening science in the NATO nations. The Division of Scientific Affairs, which has a small international staff of scientists and administrators, implements the decisions of the Science Committee on a day-to-day basis. The three major programs under the cognizance of the Science Committee are (1) the Science Fellowships Program, (2) the Advanced Study Institutes Program, and (3) the Research Grants Program. Under the first, scientists from NATO countries are given the opportunity to study and work in countries other than their own, with particular consideration given to scientists from the less developed NATO countries who are enabled to engage in activities for which their own countries do not provide the necessary facilities. Conversely, scientists from the stronger countries are given an opportunity to work in some of the less developed countries and to assist in the training of students who otherwise would not be able to receive specialized instruction in their chosen fields. The current budget for this activity is approximately $ 2-1/2 million. Under the second program, scientists may arrange conferences at which specific topics are discussed at an advanced level. These meetings, which have various formats, may range from highly specialized colloquia to summer schools for graduate students. They vary in duration from one to eight weeks and permit exhaustive treatment of a given scientific topic by men whose reputation is world-wide. Participation is held to a sufficiently small number, approximately between 50 and 100 participants for each study group, so that active participation from all members can be expected. About 40 summer institutes are planned for the current year, and about 1,000 scientists are expected to participate. The annual budget for this program is approximately $0.5 million.

Under the third program, the Division of Scientific Affairs supports with grants a number of research workers in various countries, particularly in those countries where national resources are insufficient. A large part of the support is reserved for projects which require international cooperation and so enhance the function of NATO as a cooperative institution. About $1 million per year is available for this part of the program.

Other activities of the Division of Scientific Affairs and the Science Committee include the organization of advisory groups in specific fields where international cooperation is essential, such as oceanography and meteorology or in fields where an intense effort is deemed to be appropriate, such as human factors and operational research. One of the major activities of the Science Committee was a study under the title "Increasing the Effectiveness of Western Science," which was undertaken in cooperation with the Ford Foundation and the Fondation Universitaire of Brussels. The members of this study group included several of the most distinguished scientists from NATO nations, and the report published by the Fondation Universitaire has received a great deal of attention.

In general, the Science Committee and the Division of Scientific Affairs have devoted more attention to the general field of basic research than to specific defense matters, but the Division also maintains close contacts with other NATO activities in which defense matters play a more prominent part. Thus the Assistant Secretary General in conjunction with the Assistant Secretary General for Production, Logistics and Infrastructure convenes the defense research directors of the various NATO countries at regular intervals for technical meetings at which representatives of the Supreme Commanders and the directors of NATO research agencies and technical centers participate. The Assistant Secretary General for Scientific Affairs also maintains very close relations with the Advisory Group for Aeronautical Research and Development, the SHAPE Air Defense Technical Center in The Hague, and the SACLANT Anti-Submarine Warfare Research Center in LaSpezia.

THE ARMAMENTS COMMITTEE AND THE PRODUCTION, LOGISITICS AND INFRASTRUCTURE DIVISION

The relationship of the Armaments Committee to the Production Division is similar to that of the Science Committee to the Division of Scientific Affairs. One of the subordinate units of the Armaments Committee is the NATO Naval Steering Group which meets about three to four times per year to consider armaments for ships, ASW, air defense, and standardization of NATO equipment. From time to time the Steering Group requests the Armaments Committee to set up mixed working groups consisting of scientists, engineers, and military planners to study specific problems and to report back to the Armaments Committee and the Assistant Secretary for Production, Logistics and Infrastructure. Although these studies are mainly concerned with hardware development, they include a certain amount of research and are mentioned here only to complete the picture. Other responsibilities of the Division are to organize the exchange of technical information concerning research, improvement, and production, and to assure technical and financial supervision of the Infrastructure Program. This program includes planning, construction, and common financing of fixed installations such as airfields, telecommunications networks, fuel pipelines, etc., without which modern armies cannot operate.

THE ADVISORY GROUP FOR AERONAUTICAL RESEARCH AND DEVELOPMENT

Historically speaking, the Advisory Group for Aeronautical Research and Development, generally known as AGARD, is the oldest NATO activity in the field of research and development. Its establishment was recommended to the Council in 1951 by a conference of NATO Research Directors under the chairmanship of the late Prof. Theodore von Karman. The recommendations were accepted by the Standing Group and AGARD held its first meeting in Paris in 1952.

The governing board of AGARD consists of national delegates which are appointed by each of the NATO nations. This board meets approximately twice a year, and authority to act between meetings is delegated to an Executive Committee consisting of the AGARD chairman, four national delegates, the Director of AGARD, the Assistant Secretary General of NATO, and the Science Adviser to SACEUR.

AGARD is attached to the Standing Group of the NATO Military Committee in an advisory capacity. Its purpose is to bring together the leading aeronautical personalities of the NATO nations in order to recommend effective ways to use their research and development personnel and facilities for the common benefit of the NATO community. A small permanent staff is in charge of the implementation of the recommendations of the governing board or the Executive Committee.

AGARD does not operate any research laboratories of its own but confines its activities to the stimulation of nations. Its technical activity is conducted by panels working on selected research and development areas and through working groups organized for specific purposes. Where appropriate, individual consultants are also employed, and special projects may be undertaken from time to time. For such studies, AGARD can draw upon the aid of several hundred scientists affiliated with it as national delegates or panel members. An example is a Long Range Study on Military Requirements which has recently been completed under the leadership of Prof. von Karman. AGARD is also engaged in Operational Research activities in coordination with the NATO Division of Scientific Affairs.

The main objectives of the Panels are:

(1) To facilitate the interchange of research and development information;
(2) To indicate gaps in present knowledge and advise on how they should be filled, and
(3) To coordinate research and advise on special problems.

The panels currently active are the following:

The Aerospace Medical Panel

Engineering advances have made it possible to design aircraft whose performance put a severe strain on the pilot. One of the major functions of the panel is to determine the limits of stress to which the human operator can be exposed without severely affecting his ability to carry out his task and in more severe cases, even to survive. At the same time, the panel stimulates research activities which would aid the human pilot to cope with this extraordinary environment. It has arranged a number of symposia on various subjects, such as selection methods and criteria for air crew personnel, adaption of the airplane to the pilot, aeromedical aspects of flying safety, human engineering aspects of flight in unconventional aircraft, vision problems in space and in conventional aviation, and others. In order to reach an international audience, several groups of collected papers presented during these symposia have been published and widely distributed.

Avionics Panel

This panel arranges regularly meetings and symposia on subject devoted to specific item on intense current interests. Past topics have included long-range detection of fast-moving bodies at high altitudes, data handling for air traffic control, radio navigation for airborne and space vehicles, microminiaturization of electronic circuitry, and low noise electronics. The proceedings of the panel meetings are published and made available to engineers and scientists in all NATO countries.

An Ionospheric Research Committee operating under the Avionics Panel has dealt with such subjects as radio wave absorption, disturbances of solar origin on communications, forward scatter communication with via meteor trails, and properties of the upper atmosphere.

Combustion and Propulsion Panel

This is one of the most active panels of AGARD, mainly through the organization of technical meetings for the discussion of problems of current interest. Its annual symposia alternate with those of the Combustion Institute in order to avoid duplication and to strengthen the activities of both institutions. The next symposium prepared for July 1963 will deal with Energy Conversion and Space Propulsion.

The Flight Mechanics Panel

This panel counts among its major achievements the preparation and publication of a NATO flight test manual and the initial publication of provisional recommendations for V/STOL requirements. The purpose of the manual is to aid flight test engineers, test pilots, and instrumentation personnel of all NATO nations in the planning and attainment of a sound flight test program in order to ensure comparable results. Other activities include exchange of instrumentation between nations and provisions for making available the use of air speed calibration facilities in the United Kingdom and France to other NATO nations not equipped for this purpose.

The Fluid Dynamics Panel

This panel, formerly known as the Wind Tunnel Panel, has as its main objective the interchange of aerodynamic information, in particular that pertaining to experimental techniques. It is also prepared to lend assistance to various NATO groups by suggesting and coordinating research work in the field of fluid mechanics. It sponsors specialists' meetings on specific subjects, such as the high temperature aspects of hypersonic flow or the safety aspects of V/STOL aircraft from the aerodynamic point of view.

Structures and Materials Panel

No field connected with aeronautical research and development has as broad a scope as that of materials. One of the main objectives of the panel is to develop design data for and characteristics of the materials which are available for aircraft applications in the NATO nations. Handbooks containing these data have been published for aluminum, steels, magnesium, nickel, and titanium. A handbook dealing with materials of high-temperature and high-strength characteristics is in preparation. Other manuals prepared under the direction of the panel deal with aeroelasticity and aircraft loads. A coordinated program on runway roughness has contributed to NATO specifications for future runway construction.

Technical Information and Documentation Committee

Since the most conspicuous products of AGARD are reports (AGARDographs), memoranda, handbooks, and miscellaneous publications, of which close to 500 have been published already, the Technical Information and Documentation Committee plays a very central role. Its primary purpose is to provide service and advice in documentation for the technical panels of AGARD and to advise them on publication problems. In addition, the individual panel members exert a considerable influence on their national agencies in regard to the publication and propagation of aeronautical information. One of the major

achievements of the panel was the production of the AGARD Aeronautical Multilingual Dictionary on aeronautical terms in eight languages, namely Dutch, English, French, German, Italian, Russian, Spanish and Turkish. For each term, a definition of its concept is provided in all of the eight languages mentioned. Other activities deal with the standardization of technical reports, the provision of scientific and technical information on index cards, the preparation of specialized bibliographies, including a recent literature survey on magneto-fluid dynamics, the publication of a report, "Sources of Translations of Aeronautical Interest," and many other phases of documentation and information services.

International Consultant and Exchange Program

With aeronautical research facilities and technical personnel distributed very unevenly among the NATO nations, the Consultants and Exchange Program plays an important part in providing assistance to various countries through the exchange of technical personnel. Under this program, AGARD finances visits by specialists from one country to others for the purpose of giving lectures on fundamentally important problems, for consultation on equipment designs and on problems arising in connection with the establishment of new laboratories. AGARD also makes arrangements under this program for the exchange of research instruments, or loan of such instruments to countries wishing to carry out specific research programs, or for the organization of short duration courses enabling engineers from NATO nations to get acquainted with specific techniques. Activities are initiated either by request from national delegates or by recommendations of the AGARD panels.

TRAINING CENTER FOR EXPERIMENTAL AERODYNAMICS

There are few nations, and hardly any institutions of higher learning, which possess all the research and educational tools required for the training of aerodynamicists in modern techniques. The main purpose of the Center is to make available under one roof various educational models of modern aerodynamic facilities which are rarely found in a single academic institution. In this way the student will not only become familiar with one type of research equipment, but will get a basic understanding of many types that are used in modern aerodynamics. More specifically, the Center, located at Rhode St. Genese near Brussels, has the following aims:

(1) To give scientists and engineers from the NATO countries specialized training in the techniques and practices of a modern aerodynamics laboratory;

(2) To contribute to the dissemination of knowledge in the field of fluid mechanics, and more particularly of experimental and applied aerodynamics;

(3) To conduct and encourage research in the fields of experimental aerodynamics and fluid mechanics.

The Center accepts students whose preparation is equivalent to an M.S. degree of a university of the United States. Candidates must have a fluent knowledge of either English or French. The course is planned for an academic year of nine months and leads to a diploma. A number of short (three-month) courses are also offered. The Center was established as a NATO facility as the result of a recommendation from an ad hoc group of AGARD consultants. It is administered by an independent board of directors; however, the AGARD national delegates nominate students and the director of the International Consultants and Exchange Program of AGARD is an ex officio member of the Board.

THE SHAPE AIR DEFENSE TECHNICAL CENTER

To provide scientific and technical advice and assistance to the Supreme Allied Commander, Europe (SACEUR), and his staff in the field of air defense, the North Atlantic Council authorized in 1955 the organization of the SHAPE* Air Defense Technical Center (SADTC) at the Hague, The Netherlands. Its mission includes "to undertake research, studies, investigations, development projects and operational tests concerned with the air defense of Allied Command, Europe." As air defense cannot be considered completely in isolation from other military problems, these terms of reference are interpreted in a broad sense and the Center may engage in research development, and studies relating to more general applications of military technology. The charter also permits the Center to render scientific and technical assistance to individual NATO nations under certain conditions.

The Center is located on land furnished rent-free by the Netherlands government. Security services are also provided by this government.

Originally funded by the U.S. Department of Defense and operated under a contract, the Center has been transferred as of 1 March 1963 to NATO funding and administration. The present organizational structure and the command and administrative channels to the North Atlantic Council are shown in Fig. 2. The civilian director of the Center reports to SACEUR via the Assistant to the Chief of Staff. The charter also provides for a scientific committee of national representatives whose duty is to assist SACEUR in establishing the Center's work program and to provide the Center with scientific and technical advice in the field of air defense. Other duties of this committee are:

* The term SHAPE (Supreme Headquarters, Allied Powers, Europe) is used interchangeably with SACEUR.

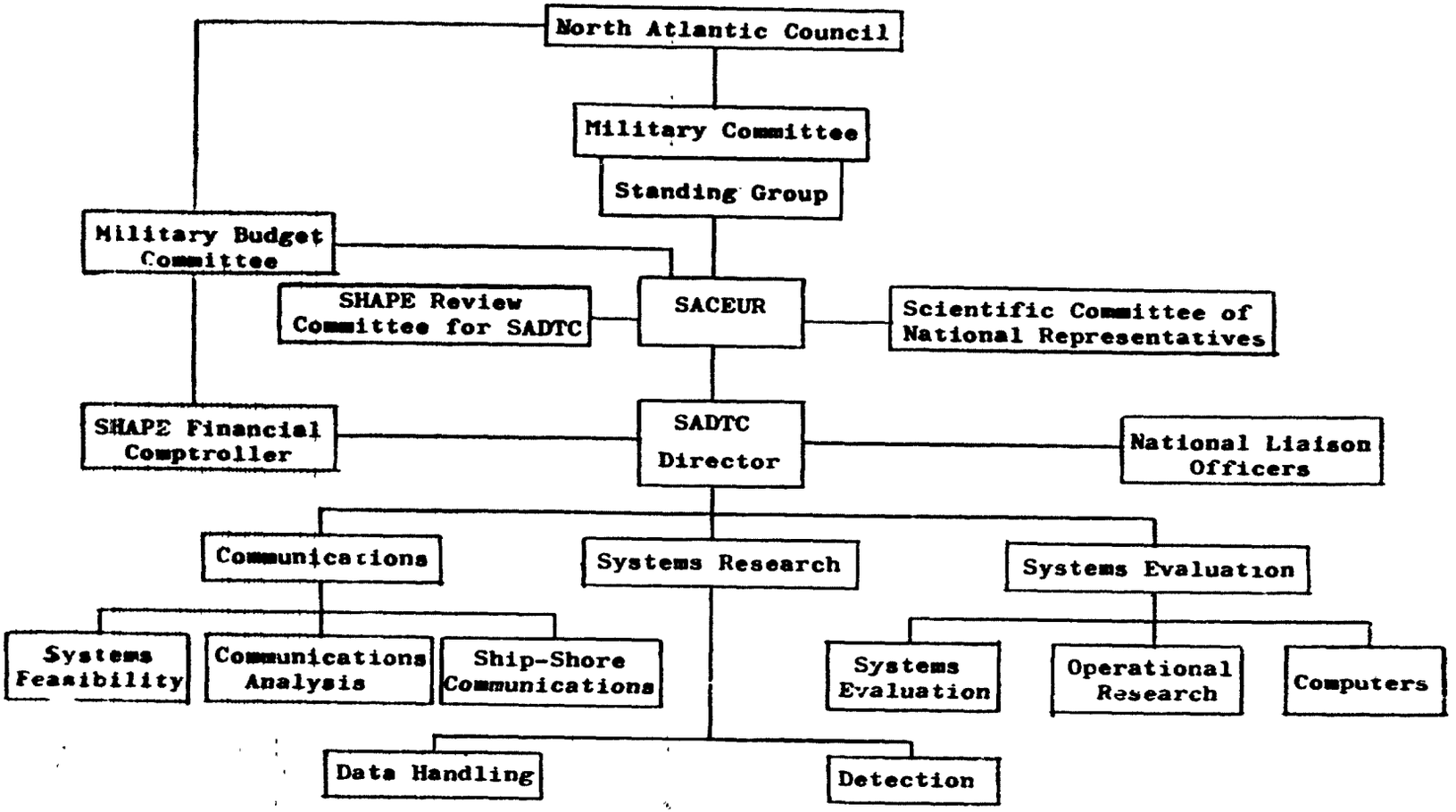

North Atlantic Council
Military Committee
Standing Group
Military Budget Committee
SHAPE Review Committee for SADTC
SACEUR
Scientific Committee of National Representatives
SHAPE Financial Comptroller
SADTC Director
National Liaison Officers
Communications
Systems Research
Systems Evaluation
Systems Feasibility
Communications Analysis
Ship-Shore Communications
Systems Evaluation
Operational Research
Computers
Data Handling
Detection

Figure 2

(1) To advise SACEUR on the selection of a director and his deputy;

(2) To see to it that all information in possession of NATO nations relevant to the mission of the Center is made available to the maximum extent acceptable to those nations;

(3) To assist the Center in the recruitment of qualified scientific personnel, and

(4) To support the Center in their own countries in every respect.

Each nation may also appoint a liaison officer to maintain regular contact with the Director of the Center. These national liaison officers will normally attend the meetings of the Scientific Committee. All NATO nations, with the exception of Iceland, have availed themselves of the privilege to appoint a liaison officer. There is also a liaison officer of the Center in permanent residence at SHAPE.

The current and projected work programs of the Center as developed by the Director are reviewed by the Scientific Committee of National Representatives and submitted to SACEUR via the SHAPE Review Board for SADTC which consists of the heads of all the working divisions in SHAPE. SACEUR will then forward these recommendations to the North Atlantic Council for approval in principle. The budget of the Center is submitted to the Council via the Military Budget Committee for approval.

The scientific program of the Center originates in three ways: Proposals for specific investigations are submitted by individual NATO nations, by SHAPE, or generated by personnel of the Center itself.

The scientific staff consists of 86 scientists, 120 supporting technicians, and 140 administrative, custodial, and security personnel recruited from 13 of the 15 NATO nations. It is organized in three major groups: Communications, Systems Research, and Systems Evaluation, and includes topics such as detection and identification, military communications, computers and data handling, and operational research. Part of the work is carried out under contract and a certain number of scientific personnel are provided on a loan basis to the Center by various national organizations and contractors.

THE SCIENTIFIC ADVISER TO SACEUR

Inasmuch as the SADTC is located in the Hague and because its mission is restricted to air defense, it was felt desirable to have at SHAPE a small scientific group which could maintain day-to-day contact with SACEUR and his military staff. This group, consisting at present of four scientists, is engaged in studies of various military problems, that arise either from the officers at

SHAPE or are generated by the scientists themselves on the basis of their own observations. This group has no administrative responsibility for the operations of the SADTC and acts in a strictly advisory capacity.

The Scientific Adviser is an ex-officio member of the Executive Committee of AGARD and maintains more informal liaison with the NATO Division of Scientific Affairs and the Science Committee.

THE SACLANT ASW RESEARCH CENTER

The latest addition to the group of NATO research and development activities is the SACLANT ASW Research Center (SASWREC) at LaSpezia, Italy. It was founded in May 1959 to provide scientific support in the European area for the ASW mission of the Supreme Allied Commander, Atlantic. Nine NATO nations (Canada, Denmark, France, Germany, Italy, The Netherlands, Norway, United Kingdom and the United States) are currently represented on the scientific staff of the Center.

The basic reason for establishing this Center was a desire to assemble in a European location a group of scientists from NATO nations to bring together new viewpoints and a fresh look at critical scientific problems of ASW which have thus far resisted successful solution. The primary mission of the Center, therefore, is to conduct scientific research programs on topics of importance to anti-submarine warfare. In addition, it provides technical advice and consultant services to SACLANT and through him, to other NATO Commanders. It is not, however, the purpose of the Center to develop specific pieces of hardware.

The Center occupies a laboratory building, provided rent-free by the Italian government, located in an Italian naval reservation at the Gulf of LaSpezia about four miles from the center of the town. The Italian government also provides custodial and security services for the Center.

The facilities of the Center consist of a set of electronic and oceanographic laboratories, workshops, a technical library, a high-speed digital computer installation, and the research ship ARAGONESE. This ship, a converted 3,000-ton freighter has accomodations for a technical party of 12 and is equipped for oceanographic, electronic, and acoustic work at sea. For shorter excursions, the Center has a 63-foot work boat furnished by the U.S. Navy. Most of the actual research effort is carried out at sea and the laboratory work ashore consists mainly of design, construction, maintenance, and calibration of scientific equipment to be used on the research ship, and last but not least, evaluation of the data obtained during the various cruises at sea.

Since its initial commissioning in May 1959, the organization and administration of the Center has undergone a number of changes. Originally funded by the Department of Defense of the United States under the Mutual Weapons Development Program (MWDP) and operated under contract, first by a commercial concern and later by a non-profit organization, it has been transferred as of 1 February 1963 to NATO funding and administration. The present organizational structure and the command and administrative channel to the North Atlantic Council are shown in Fig. 3. It should be emphasized that the civilian director of the Center reports directly to SACLANT, who provides policy direction and approves the scope of the annual program of the Center, including the magnitude of the effort in different areas.

The charter for SASWREC, which is very similar to that for SADTC, provides for a Scientific Committee of National Representatives which has the same duties in respect to SACLANT as the corresponding committee in the case of SADTC. Each participating NATO nation may also appoint a liaison officer to maintain regular contact with the Director of the Center. These national liaison officers are supposed to inject the operational viewpoint into the program; they will normally attend the meetings of the scientific committee.

The Center obtains advice on administrative and personnel matters from a representative designated by the Secretary General of NATO (at present this post is held by the Assistant Secretary General for Scientific Affairs) and on military matters from a U.S. naval officer named as Naval Adviser to the Director.

The scientific program of the Center is carried out by five groups: Sound propagation, target classification, non-acoustic ASW, oceanography, and operations research. The scientific staff consists of approximately 36 scientists and a slightly larger number of supporting technicians. The largest group is primarily concerned with the propagation of sound in the ocean, with special emphasis on the acoustic path as reflected from the surface and bottom. Other groups are working on classification of submarine targets by sonar techniques, and finally with the exploration of non-acoustic methods for submarine detection.

The oceanographic group, consisting at present of eight scientists and an appropriate number of technicians, is engaged in research on two program areas. The sub-group on physical oceanography is concerned with the structure of the ocean bottom and its effect on acoustical problems and with such problems of submarine geophysics as the recording of gravitational and magnetic data, including measurements of geomagnetic fluctuations. The sub-group on military oceanography deals with relations between oceanographic factors and the performance of ship-mounted sonars.

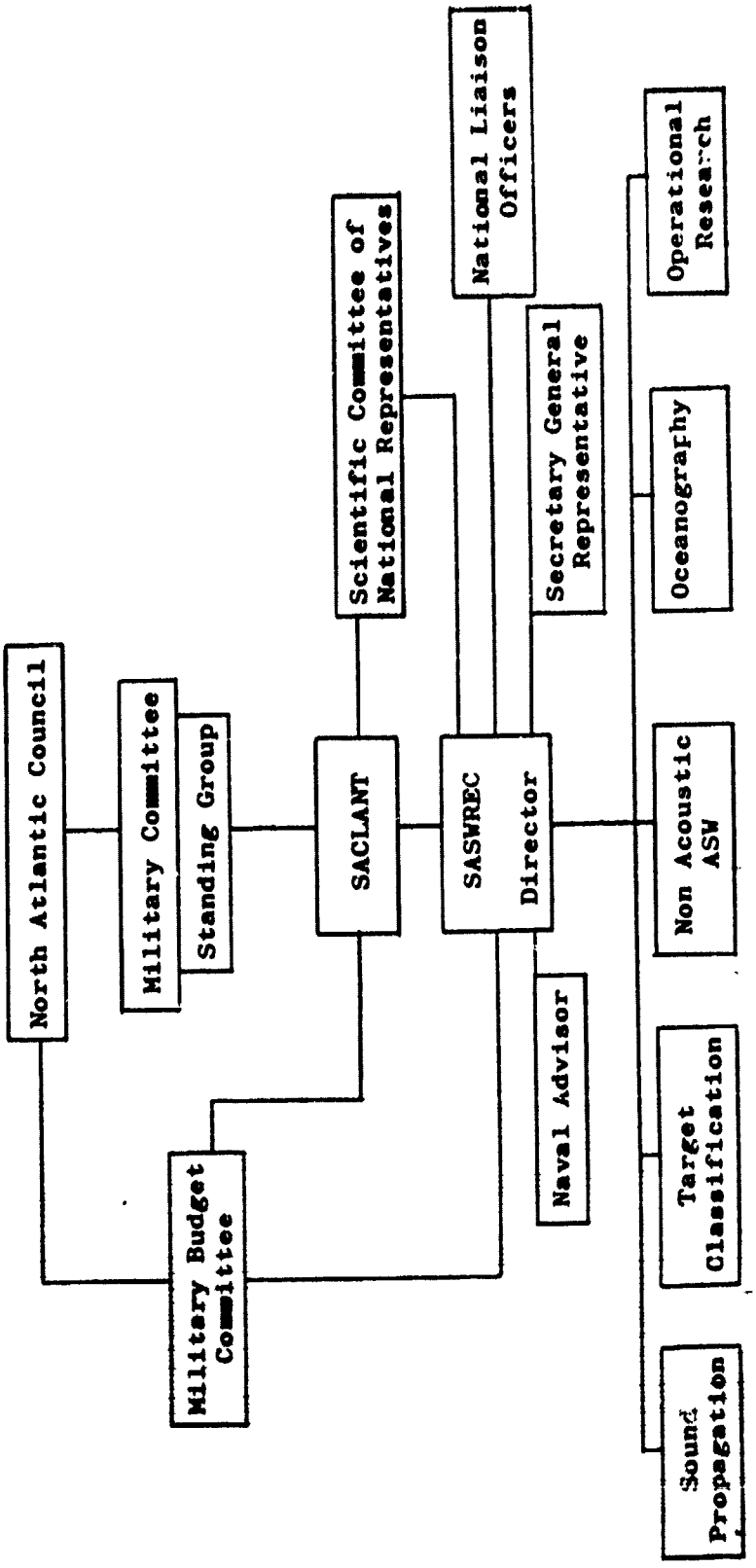

North Atlantic Council
Military Committee
Standing Group
Military Budget Committee
SACLANT
Scientific Committee of National Representatives
SASWREC
Director
National Liaison Officers
Secretary General Representative
Naval Advisor
Operational Research
Oceanography
Non Acoustic ASW
Target Classification
Sound Propagation

Figure 3

An area northwest of Corsica which provides depths up to 1400 fathoms has been selected for a very thorough investigation of all the relevant oceanographic factors in order to provide better correlation between oceanographic data and prediction of sonar ranges.

The operational research group is working on projects having particular pertinence to ASW within the NATO context, such as assessement of possible future ASW weapons, parametric studies of ASW effectiveness, and analysis of ASW fleet exercises. The group also undertakes work on specific problems as requested by SACLANT.

The scientific staff of the Center is composed of members of all the participating NATO nations while the supporting staff is mostly Italian. English, however, is the official language and is used for all reports and technical documents. The total staff now numbers approximately 150 and the annual budget approximately $2 million. A slight increase of personnel is planned for the immediate future and it is expected that the scientific staff will be built up toward a goal of 50 scientists within the next two years.

xxx

INNOVATION: THE NEW RESOURCE

Rear Admiral Lloyd V. Berkner, USNR
Graduate Research Center of the Southwest

We are assembled here in Pensacola, on behalf of our nation, with the major task to undertake a review and stocktaking of the program of research and development that supports the strength and vitality of the Naval Forces of the U. S. A.

Since World War II, our nation has learned that to maintain world peace, we must be constantly on the alert. There are nations that are ever on the lookout for any opportunity to extend their power and authority over free peoples. Only a viable American military strength in full readiness can thwart those unholy ambitions. We have learned that with any sign of weakness, the consequent rise of enemy overconfidence encourages him to embark on rash adventures that could precipitate the world in a destructive holocaust. Any advantage acquired by a potential enemy creates a highly dangerous condition that threatens the peace. Therefore we must constantly improve our weapons and tactics so that they are always competent to deal effectively with any conceivable situation. Herein lies the critical importance of our task--to ensure that the most advanced scientific thinking of our time is thoroughly explored for its military implications, and scientific opportunities adapted to maximize our strength.

We are all aware that the character of military power began a major evolution with the turn of the Twentieth Century. That evolution arose from the growing effectiveness of science, and the consequent engineering derived from it. Up to the end of the last century the rifle and the cannon, coupled with hand-to-hand combat, dominated the scene. But our century saw the beginnings of subtle change--steam power from the science of thermodynamics, aircraft from aerodynamics, and radio from electromagnetic theory.

Of course, science has always exercised an influence on military strength, from the burning glass of Archimedes at Syracuse, to gun powder derived from the earliest chemistry. Indeed, over the

ages, science has always been strongly stimulated and exploited by the human desire to acquire military advantage out of scientific discovery. Yet until the turn of this century, science was not strong enough to exercise a really major influence over military doctrine.

The mighty changes of the power of science during the 20th century have revolutionized the situation. With the turn of the century we began the acquisition of an altogether new and far more powerful science. The names of Planck, Einstein, Rutherford, Bohr, Heisenburg, de Broglie, Shrödinger, Dirac, Chadwick, Hahn and Strassman stand out in the creation of this new science. In a mere 35 years, from 1905 to 1940, we acquired the science of the miniscule, the science of the particles from which matter and energy is formed. We learned the limitations beyond which the time-honored principle of cause and effect had no further meaning. Out of these revolutionary scientific advances emerged our comprehension of the wave-structure of the universe, and the wave-mechanics that governed it. This was a science of altogether new power in its ability to control and manipulate natural processes.

Almost immediately, with World War II came the opportunity to test this new science. It gave us radar to extend our vision a hundred times. It gave us new concepts of acceleration that vastly increased velocities. It gave us energy from the nuclei of matter that provided the atomic bomb. For the first time in history the application of a scientific discovery ended a war forthwith!

Even more important, we had made the discovery that the new science was so powerful that when applied consistently to any technology, it could revolutionize that technology.

In the retrospect of two decades, we can now perceive that this discovery was the beginning of a revolution in human affairs. For this discovery was the foundation of what we are now beginning to recognize as the technological revolution. This technological revolution, as it develops toward the end of this century will doubtless be the most profound event in human history. It gives man a new command over his environment, a command we do not yet fully appreciate or indeed even fully comprehend. It is a power that can free man from drudgery, to give him opportunity for the full range of exercise of the human mind.

In these very times, in the midst of the most revolutionary scientific thought of man's whole history, we find ourselves, as a group, charged with discovery and development of the ideas from which the Navy of the future must rise. As we look back, with 20-20 hindsight, we see the incredible change in our Navy over the past half century. We know that in the next half century, to meet the enemy's challenge, an even greater change in weapons and tactics must occur. We are the men charged with the direction and initiation of that change. Are we really up to this job?

With the unfolding of the technological revolution we have learned that the day of the long haired inventor is over. No longer do the ideas underlying the rise of great industries or great systems come from the Thomas Edisons or the Alexander Graham Bells, or the Henry Fords, working from their powerful but untrained imaginations. The complex technological advances of today are derived from the very boundaries of knowledge--from the sophisticated search of advanced science for the ideas from which the important innovation of our time is now derived.

In this changed environment, as a consequence of the technological revolution, our problem is to develop a system of management and operation that can optimize the derivation and transformation of ideas out of science that are applicable to more effective tactics and supporting weapons systems.

What are the elements of such a system of management and operations?

First of all we must have intimate access to the evolving ideas of science. To do this effectively we have learned that we must actively participate in the advancement of science on all fronts.

Second, from the ideas evolved from enlarging scientific knowledge must be derived the conceptual applications. This is the major step of "invention" requiring a complete knowledge of the capabilities of science, an acute appreciation of man's potential needs, and an unlimited creative imagination in conceptually transforming basic science into something that can efficiently advance man's capabilities.

Third, from the conceptual invention follows the preliminary appraisal of the practicability of the idea. This involves realistic assessment of probable effectiveness, cost of development and of ultimate production, relative advantages in competition with existing methods, elaboration of altogether new opportunities offered, and technical problems in effectuation.

Fourth, is the applied research required to remove technological blocks standing in the way of realization. This involves directed research to solve and remove rather well defined problems.

Fifth and finally, is the actual development of prototypes for test and evaluation of hardware in the conceptual tactical environment.

These then are the elements of innovation in this technological revolution for which a system of management and operation must be perfected:

1. Access to science
2. Invention out of science
3. Appraisal
4. Removal of technical blocks
5. Development of prototypes

Our job in the Navy is to perfect that system of management and operation in our revolutionary technological environment, which can produce useful and effective innovation at an optimum rate.

Let us look just a bit more closely at these elements of innovation for which management is required.

The Office of Naval Research is the Navy's spearhead in access to science. Through contracts with, and reports from university scientists; through association with the Naval Research Laboratory and other government scientific institutions; and through contacts and contracts with the scientific laboratories of industry, the Navy maintains intimate knowledge of the whole range of our scientific advance. Certainly our methods of this aspect of our management of innovation is already skillfully developed.

Efficient management of the problem of invention is a much more difficult matter. Today the process of invention is not well understood. We know that the inventor of our time must be a highly trained and motivated scientist or engineer who on one hand has complete access to the relevant scientific knowledge, and on the other, has an imaginative vision of potential requirements and applications. One is often astonished when, after lying fallow for a long time, some potent scientific idea or principle is brought alive into a major invention by the facile mind of an imaginative inventor. Sometimes invention flows from the concentrated interaction of the brains of a combined group of scientists and engineers in a highly organized brainstorming session.

I recall participating in the invention of scatter propagation at MIT in 1950. Eight scientists and engineers came together for several weeks to re-examine the whole of electromagnetic theory and its terrestrial environment for new forms of high-frequency propagation over longer distances. Everyone knew at the start that the task was hopeless--yet in a few weeks several inventions became obvious. But they appeared only after studying and debating, and questioning every statement of theory and experiment that had been done. These inventions involved a considerable extension of theory on the spot--theory that then suggested quick confirmatory experiments. Out of this deliberate inventive effort have grown whole new communications systems and great industries.

But since we do not understand the process of invention very clearly, we do not know how to manage it very well. Perhaps an extensive research should be done to identify the circumstances and environments from which specific inventions out of science now

arise. Certainly there are plenty of examples to be studied. I realize that most people will say that invention cannot be other than a matter of sheer chance, involving the accidental conjunction of a variety of unpredictable circumstances. But my intuition, sharpened by long experience, tells me that this is not quite correct. I believe that a doctrine could be evolved to increase greatly the probability of useful invention. But at the moment, we must admit that, out of ignorance, our national management of invention is a little too much influenced by undirected and unplanned chance.

Given the invention, the succeeding steps are more straight forward. Nevertheless they require great skill and imagination in their management.

Realistic appraisal of invention involves great comprehension, an imagination guided only by technological reality, and a tremendous recourse to experience. The character of appraisal depends of course on the character of the inventive proposal. Improvements of existing weapons can, of course, be assessed by the Bureau or Office responsible for them.

But the really great inventions, with their consequent major tactical evolutions, transcend the capabilities of any single internal Naval organization. One suspects that the most powerful ideas--ideas whose synthesis will really give us our Navy of the 21st-Century--should each be appraised by a special Navy-wide task-force. For such a task force, we need highly trained and imaginative officers, scientists, and engineers from the Navy Bureaus and Offices, with representatives of their laboratories and contractors. But mere technical assessment of great conceptual military advances is not enough. We need, in the appraising task force, men from the fleet and from the war colleges who can examine the whole range of tactical advances implicit in powerful inventive innovations. In the appraisal stage, these men must work together in conceiving and conceptually testing the inevitable alternative combinations of technical forms with their emergent tactical potentials. Some responsible naval office, such as ONR, should perhaps be given the responsibility of initiating and quarterbacking broad appraisals of conceptual systems emergent from scientific discovery and consequent invention.

Doubtless, the stage of appraisal is critical. History is replete with examples of failure to appraise the real worth of a great invention--this very failure was the direct antecedent of the split between the Army and Air Force. Fortunately, the Navy has been quick in its response to invention with its early espousal of carrier-based aviation, the nuclear submarine, and the strategic and the tactical employment of the Polaris missile.

With the onset of the technological revolution, these great innovations are only portents of the mighty changes for which we must gird ourselves in the future--the Navy of the 21st Century. This new technological environment warns us that we must constantly review our posture of management so that no opportunity for greater strength is lost or deferred.

The final steps toward innovation--removal of technical blocks, and development of prototypes--is well within our capacity. Once the requirement is set, the Navy has always been capable of a superb response. The Navy's performance in creating the Polaris nuclear-submarine is perhaps the most distinguished chapter in the history of military weaponry and naval tactical development.

In assessing the spur of innovation to the improvement of our Naval Forces, we cannot ignore the consequent benefits to our total economy. Throughout history, nations have risen to greatness from the advanced development of their technologic capabilities. Our military encouragement of science and our systematic exploitation of that science for useful application exerts a powerful influence on our entire national technological posture. Out of this advanced technology, so strongly developed out of military need, arises a host of new civil opportunities for new products and services useful to man. Inevitably, these play a major role in our rising economy. So aside from our defense, our military technology provides a high level of national technological dexterity from which flows man's ever-improving capability to command his environment for his benefit. In these days, when our natural resources approach full development, innovation out of advanced science and technology becomes the new resource on which our national economy will rise.

In concluding, I cannot avoid mention of one major auxiliary problem--the problem of manpower trained to command the boundaries of knowledge. In the context of the technological revolution, such men are required, at every stage of innovation, in large numbers and over the whole range of scientific fields.

Before World War II, our education at graduate levels was growing healthily. In 1920 ten great American graduate universities had emerged to produce the bulk of our doctoral graduands. By 1940, this had increased to 20 universities. Now in 1963--more than two decades later--20 universities still produce two-thirds of our Ph.D.'s. Since the war, emergence of new great graduate universities has come to a halt. During the same interval, with the onset of the technological revolution, the need for vastly increased numbers trained to the limits of knowledge has become critical. When the nation needs a quantum jump in opportunity for graduate education, we find, beyond the top twenty, the universities are faltering.

In the last decade, while the Soviet Union has increased its engineering graduates to 130,000 per year, ours have dropped from 50,000 to 10,000. Of the increased share of doctoral graduands from American universities, a substantial proportion represent students from abroad.

The situation is now so critical that every segment of our nation, private, state and national, should contribute its part to its correction. As a major employer of scientific and engineering manpower, both directly and through its contractors, the Navy cannot avoid every possible measure to stimulate education of the men needed for its own future effectiveness. With such stimulation, we can only hope that the leaders of universities 21 to 100 will awaken to their national responsibilities. Americans have risen to crises before, and I think that with the critical realization before them, they will now rise to this present crisis in advanced education. As the President's Science Advisory Committee points out--we must double graduate opportunity by the end of the decade.

Looking ahead, the future of our Navy is bright. We have been quick to establish, through ONR, an unparalleled access to science. We have been successful in retaining world leadership in naval weapons and tactics. As we face the full power of the technological revolution, we search actively for those methods of management and operations that can turn this revolution in technology to our best advantage through wise employment of innovation.

Commanding the seas that cover three-fourths of the Earth and provide access to most of it, with unparalleled mobility and readiness, and with an almost instant capability for response wherever trouble may arise, our Navy will continue to provide that defense within which freedom can prosper and free men can live in peace.

PRESENTATION OF THE CAPTAIN ROBERT DEXTER CONRAD AWARD

Rear Admiral L. D. Coates, USN
Chief of Naval Research

The pleasant task has fallen on me to present the annual Captain Robert Dexter Conrad award for an outstanding contribution to the advancement of Naval science. This is the highest award conferred by the Office of Naval Research, and the man who receives it must compete with a number of eminent nominees. Starting with the first presentation of this award to Dr. Alan Waterman, we have been careful to select those men, civilian scientist or naval officers, who have done most to advance naval science in a bold, pioneering manner as exemplified by Captain Conrad.

As the first head of the Planning Division at the time ONR was established, Captain Conrad was the primary architect of the Navy's basic research program and an apostle of the increasingly important role of scientific research to the Navy. The policies and principles of operation of the Navy's early postwar research program are still generally in effect today.

Captain Conrad himself was heir to a long tradition of the Navy's close relationship with science that goes back to the first half of the 19th century. It was in 1830 that the Navy inaugurated its first scientific program with work in the field of astronomy, magnetism and meteorology. This program formed the nucleus of what became the Naval Observatory which opened formally in 1844. Significantly it is the scientific director of the Observatory whom we are honoring tonight.

The fact that astronomy was the Navy's earliest area of scientific study is no happenstance. The Navy's ships had to navigate the trackless expanse of the world's oceans with only the sun, moon and the stars to serve as guideposts. A key factor in navigation is precise time. Since its establishment, the Naval Observatory has had the responsibility for the determination of time in the United States.

In recent years, especially since the end of World War II, the need for more precise time has increased at a rapid rate. Especially for missile and satellite tracking it has become necessary to measure time with great precision. In order to obtain this extreme precision the master clock at the Observatory, which controls the Time Service Broadcasts, is an atomic clock. Since January 1, 1958, the Naval Observatory has been keeping "atomic time" by averaging its atomic clock output with eight others distributed around the world. All nine of these clocks detect and amplify the electromagnetic waves radiated by a beam of cesium atoms passing through a tuned radio-frequency cavity.

Some of the most important data on which calculations are based are the movements of the sun and the planets, and it is in this area that Dr. Gerald Clemence has made his greatest contribution. Early in his career Dr. Clemence initiated and conducted a long series of research studies concerning the theories and tables of motions of the sun and the planets. The tables and theories then in use had been devised by Navy scientists in the late 19th century and had been used for half a century to fulfill the astronomical and navigational needs of that time.

However, advancing technology and the more complex equipment under development just prior to World War II made demands that showed that the tables then in use were beginning to deviate seriously from observation. Some investigations had indicated where temporary adjustments and corrections could be made. However, the rapid advances being made in other areas required far more accurate and precise predictions.

Dr. Clemence's work in obtaining much of the required data through his studies in celestial mechanics won him national and international recognition. His most significant studies delineated the motions of the planet Mercury based on a careful study of all the astronomical observations of that planet recorded between 1765 and 1937. The data presented in these studies were developed with such precision that they form the best proof currently available of the validity of the theory of relativity. Using a technique unprecedented for such an analysis, he corrected the observed positions of Mercury for errors introduced by retardations and fluctuations in the earth's rate of rotation.

In addition, Dr. Clemence first proposed a specific method for calculating a uniformly varying time generally known as Ephemeris time. Both the concept of Ephemeris time and Dr. Clemence's method for calculating it were adopted in 1960 by the International Astronomical Union. The determination of Ephemeris time is now a daily task at the Naval Observatory.

Another of Dr. Clemence's achievements was his development and administration of a program for investigation of the motions of the planets and satellites. This program was organized in 1947 in cooperation with Yale University Observatory and the Watson Scientific Computing Laboratory, with the support of ONR. The purpose of the program is to overcome inadequacies of current theories of motion of the principal members of the solar system.

In this connection, Dr. Clemence has organized long-range programs for systematic and complete revision of the theories of the orbital motion of the planets about the sun, of the moon about the earth, and the rotation of the earth about its axis. Results of this program have been published in many papers.

Dr. Clemence has been employed by the Navy for more than 30 years. His research programs have not only provided the Navy with a vastly improved capability in the fields of astronomy and navigation but have proven of great value both to our allies and to the scientific professions. Under his direction, the U. S. Naval Observatory has grown immensely in stature and reputation.

I am, therefore, pleased to present to Dr. Gerald M. Clemence the seventh annual Captain Robert Dexter Conrad award.

(After Presentation)

The citation reads as follows: "For outstanding contributions to celestial mechanics theory for navigational and space technology purposes. Being foremost in your field, the foresight, wisdom and industry which you brought to bear have greatly assisted the Navy and this nation.

"As Scientific Director of the Naval Observatory and consultant to the Chief of Naval Operations and Chief of Naval Research, your leadership as a scientist and administrator has helped to make possible the high prestige with which the United States Navy and the Naval Observatory are held throughout the world."

xxxx

ACCEPTANCE OF THE CONRAD AWARD

Dr. Gerald M. Clemence
U. S. Naval Observatory
Washington, D. C.

Admiral Coates, ladies and gentlemen: I think I hardly need tell you that this event is the most gratifying one of my professional career. It is sweeter to be honored at home, among friends and neighbors, than elsewhere among strangers. It is better to be honored by one's employer than by those who sometimes seek favors in return. And it is a special delight to be honored above one's aspirations; as I have observed the eminence of the previous recipients of the Conrad Award, I have felt it to be as far beyond my reach as the moon is.

That much was pretty straightforward. It is more difficult to decide how to use my remaining time. Clearly I should not speak about anything in which I have any real competence, for you would all be asleep in no time, and I should try at least to keep you awake until Admiral Berkner begins. After much hesitation I have decided to talk about the principles that have guided me in my career with the Navy.

When I joined the Naval Observatory, fresh out of school, in 1930, I made two rules for myself:

1. To act at all times in such a way as to please my immediate superior.

2. To adopt a policy of non-aggression toward all my associates.

After fifteen years of living with these two rules I found myself the senior civilian at the Observatory. It then became evident that some modification of the rules was necessary. I did intend to please my commanding officer, but it was clear that I could not expect much help from him in conducting the scientific work. So I decided to consult the taxpayers, and since it is very dangerous to seek advice without intending to take any of it, I

chose my advisors with care--I selected the half-dozen senior astronomers in the United States. I found them all eager to help the Observatory, and most of what I have accomplished since is the result of the advice they gave me then, and many times afterwards. Incidentally, no consulting fee was ever paid for any of the advice.

I also relaxed very slightly my policy of non-aggression, to the extent of helping a few of my associates to find positions better fitted to their ability and temperament than the ones they were occupying.

Respecting military-civilian relations I also had (and have) two rules. I asked the astronomer I relieved in 1945, whom I knew to be a wise man, whether he could give me any advice about my new job (expecting, of course, that he might propose a piece or two of scientific work). After two days he gave me two rules:

1. Have no secrets from your commanding officer.
2. Before you quarrel with him, make sure the subject is really important.

I have followed these rules religiously, and I have never had a quarrel with my commanding officer, and I commend them highly.

So much for my principles as I see them. What about the truth of my story? Have I been picturing myself as I wish to be seen, or have I been picturing myself as others see me? There is no way for me to find out directly. If I ask the persons who really know me, they will say it is a true picture, whether in fact it is or not. But the truth could be ascertained if you will ask those who know me, and if you do I hope you will report back to me.

INTERMETALLIC COMPOUND BASED MATERIALS FOR STRUCTURAL APPLICATIONS

W. J. Buehler
U. S. Naval Ordnance Laboratory
Silver Spring, Maryland

Introduction:

In recent years there have been increasing demands made of the metallurgists and materials scientists to develop metallic materials with unique properties or combinations of properties. A partial list of such properties or combinations of properties might include the quality of being non-magnetic, having a low density, possessing high strength at both room temperature and elevated temperatures, having resistance to oxidation, abrasion, and corrosion, along with such specialty properties as mechanical vibration damping over a wide range of stress and temperature. These are but a few of the requirements demanded by our fast moving technology that will, in the foreseeable future, allow men to move freely in outer space as well as in the vast depths of the ocean floor, or still further travel at high rates of speed over a water surface. Accomplishing these ends is a cooperative venture that requires among other areas of technology new metallic materials with properties superior to presently available materials.

To meet these demands within the naval establishment, the metallurgists at the U. S. Naval Ordnance Laboratory, White Oak, Maryland have been conducting research on Intermetallic Compound Base materials. By definition given in the Metals Handbook (ASM), 8th edition,[1] an Intermetallic Compound is "An intermediate phase of an alloy system, having a narrow range of homogeneity and relatively simple stoichiometric proportions, in which the nature of the atomic binding can vary from metallic to ionic." In order to further describe an intermetallic compound, two typical

equilibrium diagrams which include representative types of compounds are given in Figure 1. The equilibrium diagram (or constitution diagram) is one of the metallurgist's principal tools in studying phase equilibria existing in the solid state in a two or more component metal system. Observing the diagrams given it can be seen that the hypothetical compounds given are, as the definition indicates, intermediate phase(s) in an alloy system having a relatively narrow range of homogeneity. In addition, it should be pointed out that the compound can exhibit congruent or incongruent melting (C_ND_O) and in many cases a compound may have a melting point higher than either of the two component metals from which it is formed. This can be seen in the case of compound C_LD_M in the right hand diagram of Figure 1.

Intermetallic compounds have long been recognized for their unusual physical and mechanical properties. Some compounds, particularly those composed of Group III and V metals, act as semiconductors, others resist losing their strength to high homologous temperatures. However, in spite of their versatility, the usefulness of intermetallic compounds for structural application has been severely limited by room temperature brittleness. As a result, their greatest use to date has been as a minor strengthening constituent in a ductile matrix metal or alloy.

In spite of the apparent limitations of intermetallic compounds as engineering materials, the scientific demands of the U. S. Navy required a re-examination of this heretofore unpromising area. During the course of the early research on intermetallic compounds in general, one compound containing equi-atomic quantities of titanium and nickel-TiNi-showed unique properties which included marked ductility at room temperature. It is this compound and nickel-rich variations of this compound that will be discussed in some detail in this paper.

Ti-Ni Phase Equilibria

At the inception of the present investigation on the compound TiNi and associated TiNi-base alloys, it was recognized that the constitution of the titanium-nickel alloy system in the TiNi phase area was uncertain. Actually, there were two versions in the literature[2],[3] and still another version being readied for publication.[4] This uncertainty in the TiNi composition area assisted in stimulating the metallurgists at the U. S. Naval Ordnance Laboratory to delve deeply into the alloy system.

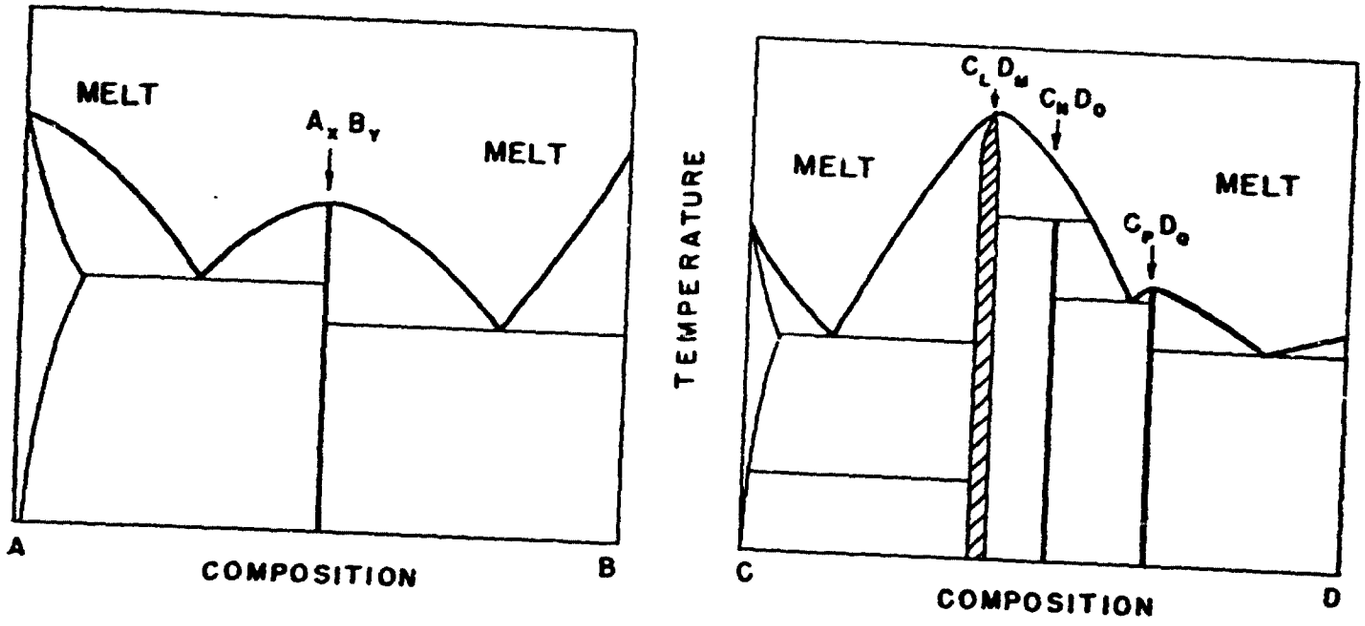


Fig. 1 Typical equilibrium diagrams showing possible ways in which intermetallic compounds may appear.

Buehler

Employing X-ray diffraction (Mo Kα radiation), metallography, internal friction measurements, magnetic measurements, and dilation studies along with many other physical and mechanical property studies, a much clearer picture of the phase equilibria has been established. In addition, it became obvious why the prior equilibrium diagram investigators had been unable to correlate their findings.

Based upon the findings at the Naval Ordnance Laboratory, a new equilibrium diagram has been proposed and is shown in Figure 2. The over-all diagram, given in Figure 2 (left), is quite similar to the work of Margolin, Ence and Nielsen[3] but this earlier investigation overlooked the "Martensitic Transformations" causing marked structural changes near room temperature. These martensitic transformations occurring in the stoichiometric TiNi compound composition range (54.5 to 55.5 w/o Ni) not only explain the confusion in the early phase equilibria investigations but also provide a combination of inherent properties in these alloys previously unreported in any other metallic systems.

Actually, in the following discussion, two titanium-nickel composition areas will be discussed; these are:

1. 54.5 to 55.5 w/o Ni* - This range blankets the TiNi compound stoichiometric composition (calculated to be 55.06 w/o Ni) and it is in these alloys that we will see the gross martensitic transformations occurring.

2. 56 to 60 w/o Ni* - This group of alloys are on the nickel-rich side and thus provides excess nickel which can be used to enhance certain mechanical properties, particularly hardness and strength.

- - - - - - - - - - - - - -

* For simplicity, the TiNi-base alloys are referred to as "Nitinol." This name is derived from the chemical symbols for nickel and titanium and the abbreviation for the U. S. Naval Ordnance Laboratory, its place of origin. When the word Nitinol is preceded by a numerical value, this number represents the nickel content of the alloy in per cent by weight. Example: 55.1-Nitinol, 60-Nitinol, etc.

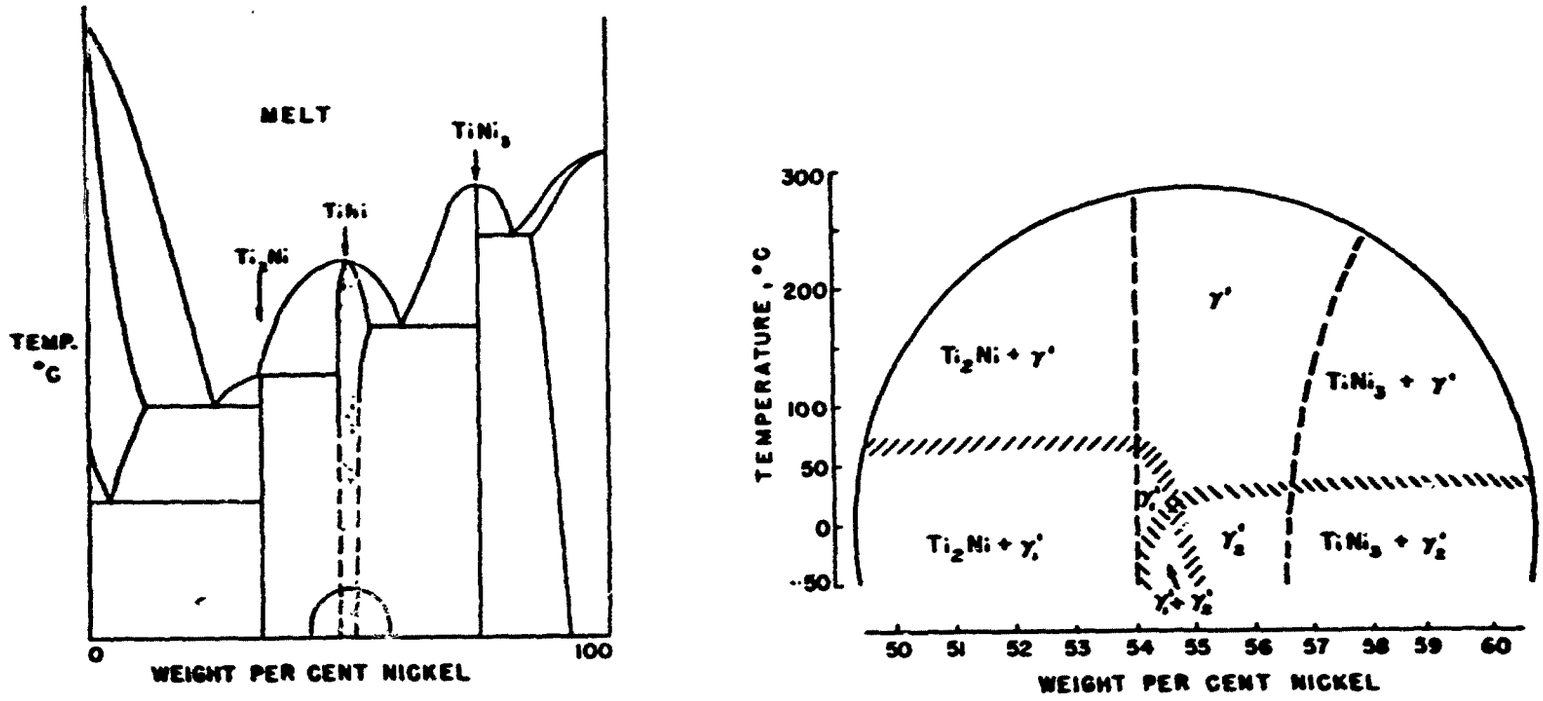


Fig. 2 Equilibrium diagram for the titanium-nickel system showing the martensitic transformations that occur near room temperature (semicircular area) in a stoichiometric TiNi alloy.

Martensitic Transformations and Related Properties

Since the stoichiometric TiNi alloys (54.5 to 55.5 Nitinol) are susceptible to martensitic transformations and these affect the resultant properties, a short discussion is devoted to this reaction. Observing Figure 2 (right panel) it can be seen that a number of phases and structures exist. These may be defined as follows:

- γ - TiNi compound phase (disordered), body centered cubic structure
- γ^1 - TiNi compound phase (ordered), body centered cubic (CsCl) structure
- Ti_2Ni - Stoichiometric Ti_2Ni compound composition, 2 atoms Ti: 1 atom Ni, complex face centered cubic structure, very brittle
- γ_1^1 - The Ti_2Ni structure, but composed of 1 atom Ti: 1 atom Ni, ductile like γ or γ^1
- $TiNi_3$ - Stoichiometric $TiNi_3$ compound composition, 1 atom Ti: 3 atoms Ni, close packed hexagonal structure, quite brittle
- γ_2^1 - The $TiNi_3$ structure, but composed of 1 atom Ti: 1 atom Ni, ductile like γ or γ^1

With the symbols and phases of Figure 2 (right panel) defined let us look at an alloy falling in the composition range from 54.5 to 55.5 w/o Ni. Upon cooling below a critical temperature (M_S), close to room temperature, a reversible structural change can occur as shown in Equation (1) below:

$$\gamma^1 \underset{\text{Heating}}{\overset{\text{Cooling}}{\rightleftharpoons}} \gamma_1^1 + \gamma_2^1 \qquad (1)$$

The structures γ_1^1 and γ_2^1 are the martensitic transformation products steming from the parent phase γ^1(TiNi) upon cooling. With heating above a critical temperature (M_D), the parent γ^1(TiNi) phase structure is reformed.

In addition to being temperature sensitive, the martensitic transformation requires some strain energy to effect the shearing process that produces the minor shifting of atoms (something less than an interatomic distance) and the resultant new atomic structure(s). In

our investigations into these stoichiometric alloys it was found that tensile stressing produced one type of martensitic product, while compressive stressing produced another. This is shown in Equations (2) and (3) below:

$$\gamma^1 \xrightarrow{\text{Tension}} \gamma_1^1 \qquad (2)$$

$$\gamma^1 \xrightarrow{\text{Compression}} \gamma_2^1 \qquad (3)$$

However, regardless of whether the martensitic phase structures are formed by temperature decrease or straining the parent phase γ^1 (TiNi) is restored upon heating above the M_D temperature.

Now that the structures have been defined and the products and stimulants for producing a martensitic transformation have been discussed, let us look at the crystallography of the transformation. Figure 3 shows, in the upper panel, a typical γ^1 (TiNi) structure through the (110) plane of the ordered body centered cubic. When the shear mechanism is initiated there is a minor movement of the atoms in an adjacent area (shown by the lower panel of Figure 3) and a new structure is produced which we call γ_2^1 and which is based upon the $TiNi_3$ structure and is close packed hexagonal. The plane between the two structures is commonly called the "habit plane." It appears conceivable in the diagrams given in Figure 3, that the $\gamma^1 \rightarrow \gamma_2^1$ transformation (or closer packing) can be accomplished by compressive force. Although the transformation from γ^1 (TiNi) $\rightarrow \gamma_1^1$ (Ti_2Ni) structure has been experimentally determined to occur without difficulty, it has been somewhat more difficult to characterize the crystallographic rearrangements that occur.

Martensitic transformations, with their minor atomic shifting and low activation energy requirements are generally accompanied by other general characteristics, a partial list of these are:

1. Martensite crystals of plate shape form on crystallographic planes.

2. Visible distortion of a polished surface is caused by the transformation.

3. Transformation occurs rapidly (10^{-4} seconds or less).

4. No composition change in the transforming regions.

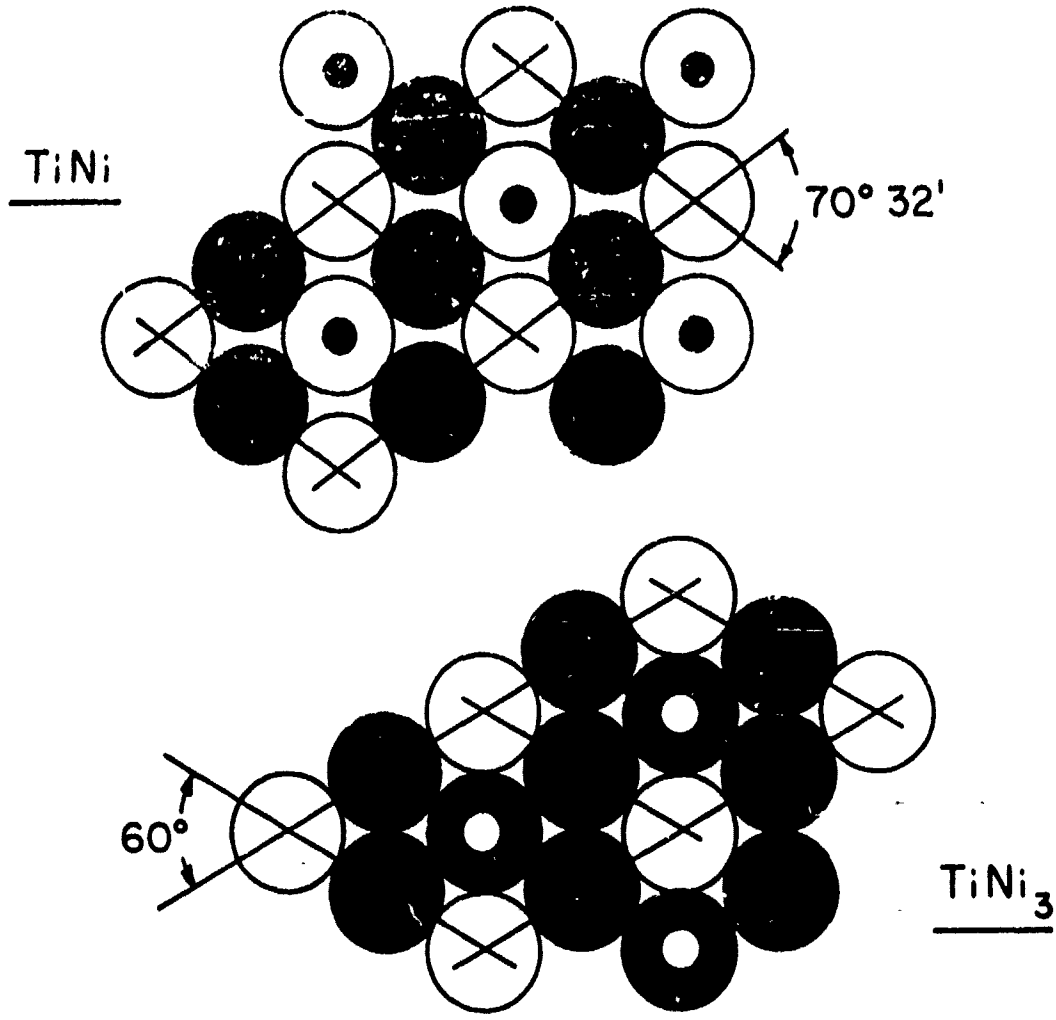


Fig. 3 Possible atomic rearrangement during martensitic transformation when TiNi (γ') changes to the $TiNi_3$ structure (ν_d). Models represent a portion of the (110) plane.

5. Low temperature martensitic phase(s) reverts back to parent phase upon heating.

Characteristics 1 and 2 above can be seen clearly in Figure 4. This is a photomicrograph taken on what was originally a polished plane surface of a stoichiometric TiNi alloy, after it had been heated to 100°C and cooled to room temperature in a "hot stage" attachment on the research metallograph. Crystals of "plate shape" can be seen both parallel and perpendicular to the original plane surface. This visible surface distortion upon heating to 100°C indicates the dimensional changes possible in an alloy undergoing decomposition into the martensitic phases or the reversion to the parent γ^1(TiNi) phase structure. A further demonstration of the surface movement during the $\gamma^1 \rightleftharpoons \gamma_1^1 + \gamma_2^1$ reaction can be seen in Figure 5. Here Knoop microhardness indentations were made in the surface of the polished specimen. Upon heating to 100°C and cooling again to room temperature, the marked reduction in the indentation size can be seen. It has been found that following several heating and cooling cycles, the indentations are almost completely eliminated.

To illustrate the speed and dimensional changes associated with the martensitic transformations, a classical demonstration is possible. In this demonstration a 55.4-Nitinol wire about 0.020 inch in diameter annealed after cold drawing is coiled at room temperature into a tight helix. At room temperature the wire remains as a tight helix indefinitely; however, if the wire is immersed in water heated above about 65°C, it reverts back to its original straight form in a fraction of a second after immersion. The reverse is also possible; a wire coiled while heated above 100°C when straightened will recoil in a fraction of a second after immersion in water heated above about 65°C.

This unique demonstration not only illustrates the speed of the martensitic transformation but it also further emphasizes the effects of strain energy in promoting the transformation. Earlier in this paper, Equations (2) and (3) showed that different martensitic phases were possible depending upon the type of loading employed. Tension forces promoting the $\gamma^1 \rightarrow \gamma_1^1$ transformation, while compression favored the $\gamma^1 \rightarrow \gamma_2^1$ change. This decomposition of the γ^1 or TiNi structure into the martensitic phases and the reversion back to the γ^1 structure with heating is shown schematically in Figure 6.

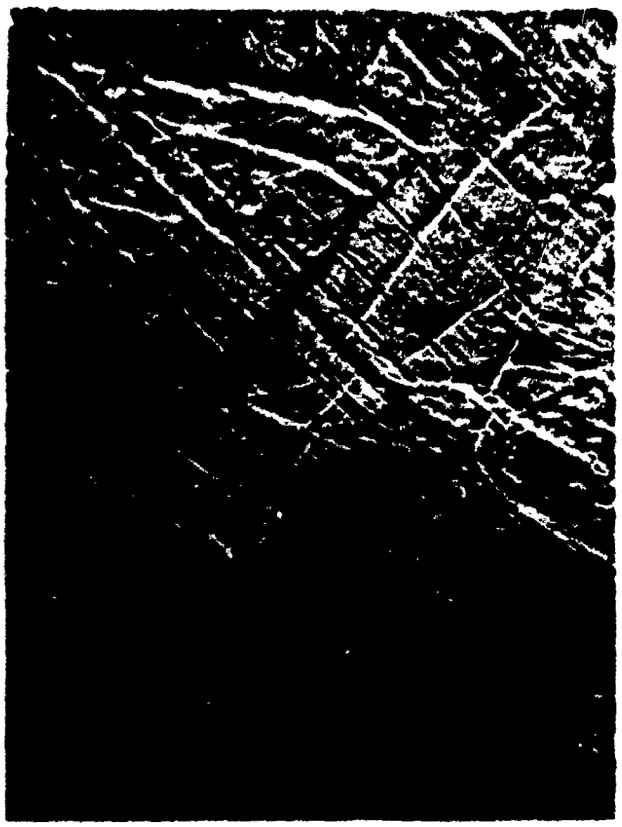


Fig. 4 Photomicrograph of a polished plane surface of a stoichiometric TiNi alloy (55.4-Nitinol) after heating to 100°C and cooling to room temperature. 500 magnifications



ROOM TEMPERATURE ROOM TEMPERATURE (AFTER 100°C)

Fig. 5 Shows drastic change in Knoop microhardness indentations caused by martensitic transformations after heating to 100°C Material: 55.4-Nitinol

How can such gross dimensional changes occur as a result of these martensitic transformations? The answer to this can best be seen by examining Table I and relating these data with the transformation mechanism seen in Figure 6. In Table I it can be seen that when there is a structural change from γ^1(TiNi) there is an associated change in the volume occupied by each atom in the new structures. The $\gamma^1 \rightarrow \gamma_1^1$ change is accompanied by a 10.63% increase in volume/atom. Conversely, the $\gamma^1 \rightarrow \gamma_2^1$ is accompanied by a reduction of 14.61% in volume/atom. By observing Figure 6 it can be seen that γ_1^1 is formed on the tension side while γ_2^1 is formed on the compression side. When heated above about 65°C, the γ_1^1 and γ_2^1 revert back to γ^1 and the volume changes act additively to straighten the specimen.

A more quantitative experiment showing the effects upon the dimensional change of different methods of plastic deformation can be seen in Figure 7. Here a given composition rod (55.1-Nitinol was parted in the middle into two equal lengths. One half was cold swaged to introduce predominantly compressive stress. The other half-length was cold drawn to introduce

TABLE I

VOLUME OCCUPIED PER ATOM IN THE THREE Ti-Ni COMPOUND STRUCTURES*

Phase Structure	Volume/Unit Atom (A^{o3}/Atom)	Per Cent Change in Volume/Atom
γ_1^1 (Ti_2Ni)	15.162	+ 10.63
γ^1 (TiNi)	13.703	---
γ_2^1 ($TiNi_3$)	11.699	- 14.61

* Based upon data given in Hansen[5]

predominantly tensile stresses. Both cold-worked specimens were made into dilation specimens and a controlled dilation study was performed at temperatures ranging up to 300°C. As was expected, in light of the prior discussion, the specimens either expanded or contracted depending upon the structural change produced during cold

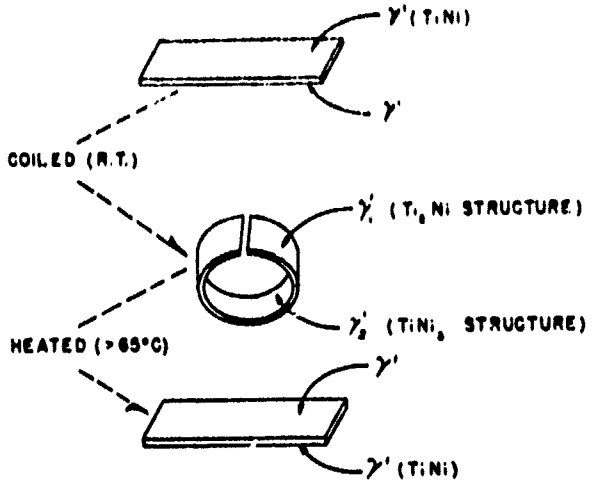


Fig. 6 Schematic drawings showing the strain induced martensitic transformations in a stoichiometric TiNi composition specimen.

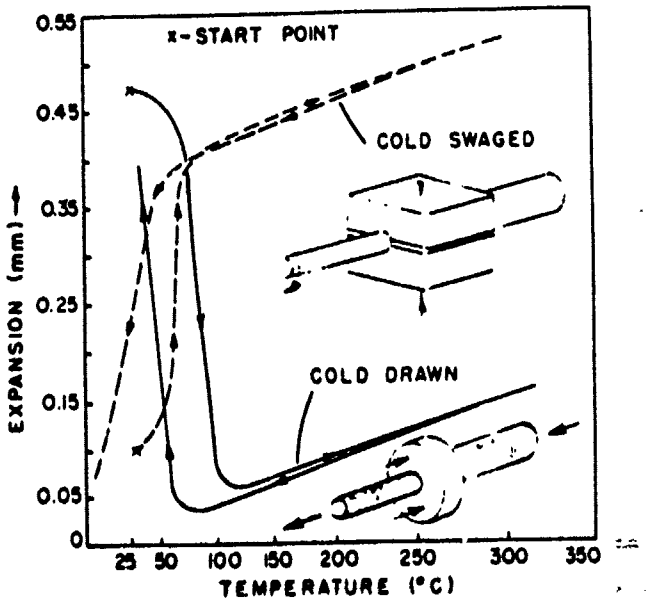


Fig. 7 Dilation curves for 55.1-Nitinol when cold deformed in compression (swaging) and in tension (drawing).

working (refer to Equations (2) and (3)), the knee in each case depicting the complete reversion back to the γ^1 or TiNi structure. Heating beyond the knee merely indicates the normal coefficient of expansion for the γ^1 (TiNi) structure. Upon cooling, the curves are essentially retraced until arriving again in the knee area; here a hysteresis occurs. This hysteresis, or failure to retrace along the heating curve, is most probably caused by the normal temperature differential between M_D (on heating) and M_S (on cooling) in the martensitic transformation.

Another example of the effect of different types of straining on the mechanical properties can be seen in Table II. Although it is quite difficult to load a specimen in pure tension, compression, etc., the effects of predominant loading are shown in Table II. From these data it can be seen that the strain induced martensitic transformations occurring during loading have the effect of producing marked changes in the elastic modulus.

Mechanical Vibration Damping

In addition to the startling dimensional changes, alloys near the TiNi composition exhibit drastic mechanical vibration damping changes as a function of temperature. Arc-cast (~5/8 inch in diameter x 4½ inches long) and hot wrought bars of similar composition and size were suspended

TABLE II ELASTIC MODULUS AS A FUNCTION OF LOADING FOR 55-NITINOL

Specimen Loading	Testing Temperature (°C)	Elastic Modulus (psi)
Tension	Room Temperature	9.3×10^6
Bend	Room Temperature	8.3×10^6
Compression	Room Temperature	4.6×10^6

from cords and struck with a hardened steel striker at room temperature and again when heated to various temperature levels. At room temperature the alloy bars containing between about 54.5 and 55.5 w/o Ni damped the mechanical vibrations like lead or tin, yet when heated to higher temperature levels the vibration attenuation went to a very low level and the bars would ring clearly. A widened curve showing the transition range from high damping-to-low damping for various composition alloys, in the stoichiometric TiNi range, is shown in Figure 8. Although the data used in plotting this curve were obtained in a somewhat qualitative manner, the results correlate well with the critical M_S and M_D temperatures in the martensitic transformation. Moving along on the almost certain assumption that the drastic damping changes are related to the martensitic transformations, a program utilizing internal friction techniques to quantitatively study this behavior is currently underway.

Explosive Behavior

Quite recently it was discovered that under certain unique conditions it was possible to cause the stoichiometric TiNi composition alloys to explode. This behavior is most unique but is again undoubtedly related to the martensitic transformations associated with this composition alloy. The only method found to date for producing the explosion is to insert small specimens of sheet material (approximately 0.040 inch thick x 3/8 inch wide) into a two-high cold rolling mill which has the rolls forced tightly together. The strip specimen is pulled into the turning mill rolls and explodes intermittently as it progresses through the rolls. The explosions that occur give off light as well as noise. The experimental test set-up for performing this investigation is shown schematically in Figure 9. The actual explosion occurs at position A, while the high velocity fragments (estimated to be 500 to 1000 feet/second) strike and scar the plexiglas shield at point B.

What causes this material to explode when drastically cold worked under the conditions imposed by the rolling mill? A possible explanation for this phenomena can be found again in the reversible martensitic transformations. A schematic drawing is shown in Figure 10 that summarizes the reaction. Observing this figure it can be seen that the strip entering the mill rolls is predominantly γ^1 or the TiNi phase; however, in most cases minor quantities of the martensitic structures γ_1^1 and γ_2^1 can exist. As the strip is rolled the pressure

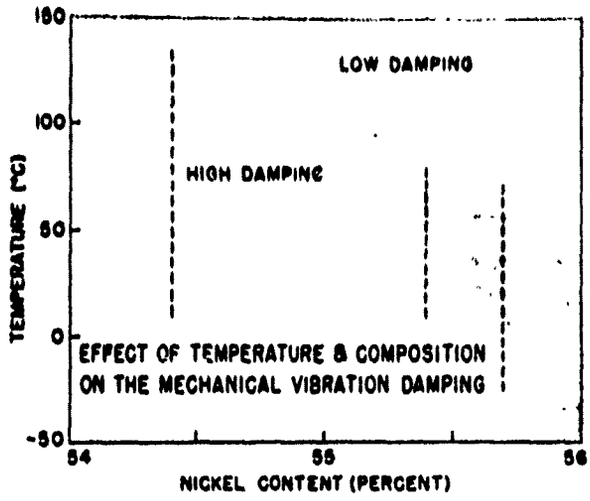


Fig. 8 Shows the damping-to-non-damping transition range as a function of temperature and composition for the Nitinol-type alloys

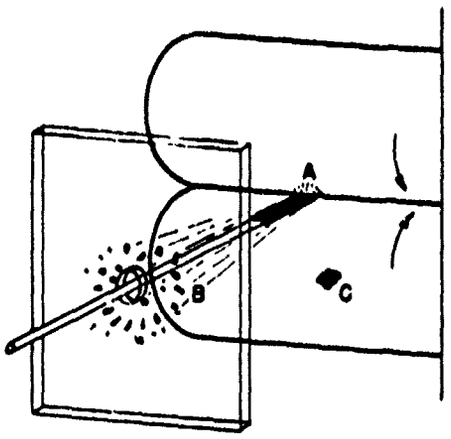


Fig. 9 Schematic drawing showing the experimental arrangement during drastic cold working. A-point at which strip enters mill. B-scarred plexiglas shield caused by fragmentation. C-Piece of spent specimen adhering to roll.

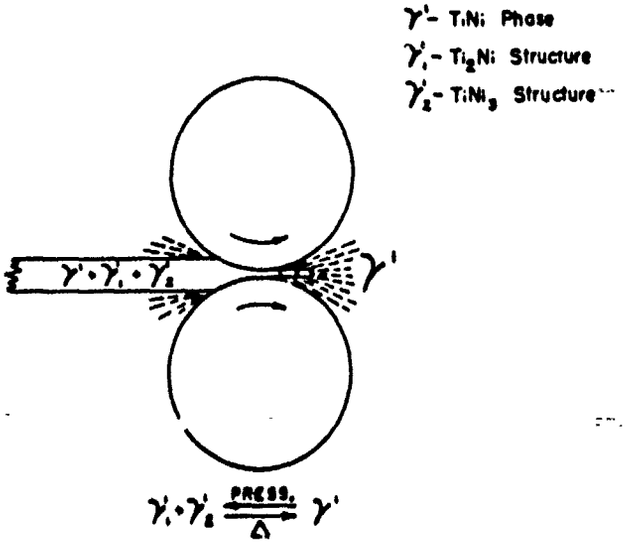


Fig. 10 Schematic drawing depicting the structural changes occurring during drastic cold rolling and the probable cause of the explosion.

promotes transformation of the γ^1 to γ_1^1 and γ_2^1, but simultaneously the strip is being heated locally by the rolling action. As the pressure is reduced (on the outlet side of the rolls), the higher temperature of the rolled material demands reversion of the unstable γ_1^1 and γ_2^1 back to γ^1. With the release of the pressure the transformation back to γ^1 occurs explosively. If the strip is heated to about 100°C prior to passing it through the mill, no explosion will occur. What this means is that the higher temperature (above the M_S or M_D) prevents the γ_1^1 and γ_2^1 phase structures from forming and results only in the drastic cold reduction of the γ^1 or TiNi phase.

Physical Properties of Stoichiometric TiNi Alloy (55.1-Nitinol)

With the existance of a martensitic transformation established in the TiNi composition alloy, and the effects of such a low temperature structural change determined, it is interesting to look at the physical properties in general. Many of the more important properties are given in Table III. From these data it can be seen that the density falls about midway between that of pure titanium and steel. The melting point is about 100°C below that of low carbon steel, and its magnetic permeability is conservatively quoted at <1.002. On this latter point the material has been found to be completely non-magnetic from almost absolute zero to 550°C, with the only possible indications of ferromagnetism coming from oxidized alloy surfaces where either nickel of the alloy combines with oxygen to form a nickel oxide or the nickel oxide formed is subsequently reduced by titanium to leave free nickel and titanium oxide (e.g., $2NiO + Ti \rightarrow TiO_2 + 2Ni$). There has been both microscopic and X-ray diffraction evidence to indicate the latter mechanism of oxidation. As a result, it is mandatory that high temperature surface oxidation products be removed from all components used in non-magnetic applications.

The electrical resistivity is somewhat similar to that of the nickel-chromium alloys and it increases almost linearly with temperature to a value of about 130 microhm-cm at 900°C. The linear coefficient of expansion and recrystallization temperature is actually very similar to that experienced in steel. The thermal conductivity, as might be expected from the fairly high electrical resistivity, is somewhat poorer than steel - actually being more like stainless steel.

TABLE III

PHYSICAL PROPERTIES OF STOICHIOMETRIC TiNi ALLOY (55.1-Nitinol)

Density	6.45 grams/cm^3
Melting Point	1310°C
Magnetic Permeability	<1.002
Crystal Structure	Body centered cubic
Electrical Resistivity	~80 Microhm-cm
Linear Coefficient of Expansion	10.4 x 10^{-6} per °C
Recrystallization Temperature	550°C - 650°C
Thermal Conductivity	Like Stainless Steel

Nickel-Rich TiNi Alloys

At the outset of this paper it was mentioned that two groups of TiNi-base alloys would be discussed. The stoichiometric TiNi composition (nominally 55-Nitinol), with its martensitic transformations and unique properties, have been previously discussed in this paper. The other area of interest is the nickel-rich TiNi composition with its fine thermal hardening capability. Figure 11 presents hardness as a function of nickel content for furnace-cooled (about 30°C per hour) and water-quenched specimens. Specimen mass was maintained fairly small to avoid the complicating factor of depth of hardening or hardenability. Observing these curves it can be seen that there is a sudden divergence of the two curves above about 56 w/o nickel. The alloys containing from 59 to 62 w/o nickel show a plateau of hardness, in the quenched condition, of about 62 Rockwell-C, while the furnace-cooled alloys of similar composition tend to maintain a constant hardness near 35 Rockwell-C. What causes this marked change in hardness with thermal treatment? Going back to the Ti-Ni equilibrium diagram (Figure 2) it can be seen that there is a sloping boundary line between the TiNi (or γ) single-phase area and the TiNi + $TiNi_3$ two-phase area lying to the right of the TiNi area. This sloping line indicates a retrograde solid solubility, or that excess quantities of nickel can be taken into solid solution in the TiNi single phase at higher temperatures. What happens when the alloy is cooled back to room temperature? If cooled slowly, the excess nickel precipitates in large globules as the second phase $TiNi_3$ (the stoichiometric compound containing 3 atoms of Ni: 1 atom of Ti).

These large aglomerates of $TiNi_3$ are ineffective in modifying the properties; as a result the hardness remains constant at a level that is characteristic of the TiNi phase matrix (see Figure 11). On the other hand, if the excess nickel is taken into solid solution by the TiNi phase at high temperatures (900°C to 1000°C) followed by a rapid cool or quench to room temperature, then the excess nickel probably forms a fine dispersion of the stoichiometric $TiNi_3$ phase that distorts the TiNi lattice and hardens the bulk of the alloy. With the wide hardness difference between the quenched and furnace cooled curves, tempering (or aging) to intermediate hardness levels has been performed successfully.

What is the significance of a high hardness through thermal treatment? The major advantage of a 60-Nitinol alloy, in addition to the increased strength associated with the higher hardness, lies in the fact that for the first time a material is available with the combination of high hardness and the qualities of being stably non-magnetic and resistant to corrosion and abrasion.

Mechanical Properties

An over-all average of some of the more important mechanical properties of both groups of alloys (nominally 55-Nitinol and the 56 to 60-Nitinol compositions) is given in Tables IV, V and VI. These tables are not included to present the entire mechanical property pictures, but rather to show the general trend in properties as related to composition and treatment.

Observing Table IV it can be seen that the ultimate tensile strength varies from about 125,000 to 180,000 depending upon the nickel content of the alloy and its treatment history. Yield strength data is not included in Table IV but it can be said that the yield strength of the 55-Nitinol will vary from 30,000 to 85,000 depending upon treatment, cold deformation and impurity content. The yield strength of the hardened Ni-rich 60-Nitinol will run close to the level of the ultimate strength. The tensile elongation, which indicates ductility, has been consistently in excess of 20% for commercially produced 55-Nitinol. The arc-cast tensile properties, of these same alloy compositions, run generally about 65% of the strengths shown for the hot wrought materials.

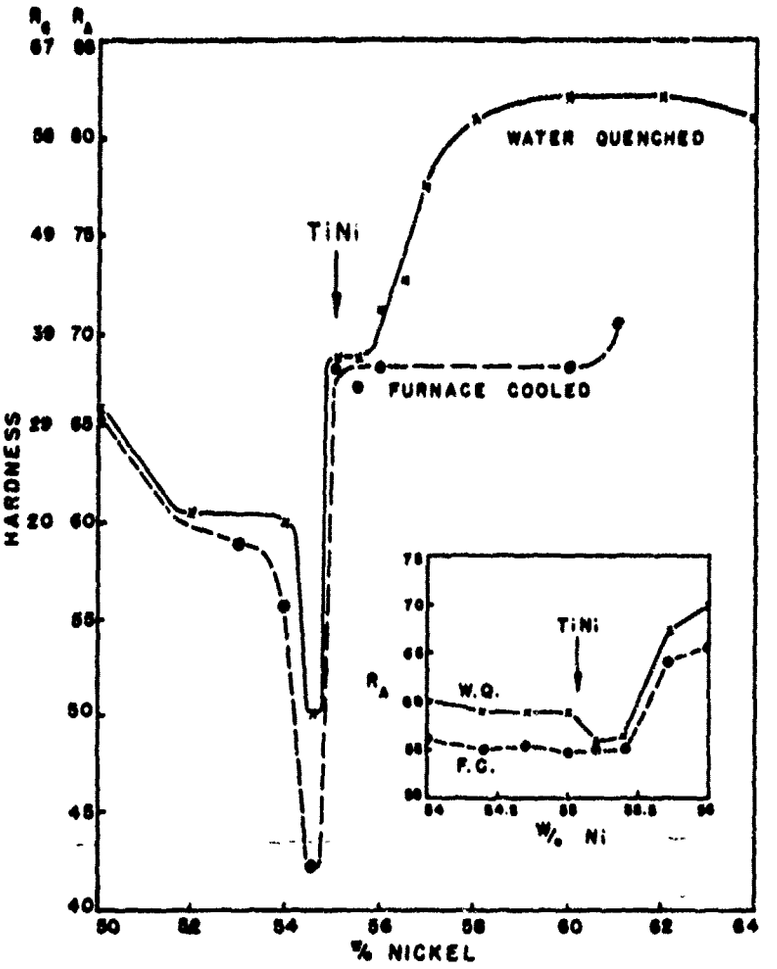


Fig. 11 Hardness curves for various composition Ti-Ni alloys (with about 0.1 w/o Fe impurity) in the furnace cooled and quenched conditions. Inset shows change in curves when Fe impurity is eliminated.

TABLE IV AVERAGE TENSILE PROPERTIES

Composition % Ni	Specimen Cond.	Specimen Hardness	Tensile Strength (lbs/in²)	Total Elong. (%)	Young's Modulus (lbs/in²)
55	Hot Worked	20 R_c	125,000	22	10×10^6
60	Extruded (hot)	50 R_c	178,000	<1	15.3×10^6
	Quenched	61 R_c	154,000	<1	16.5
	Furnace Cooled	37 R_c	137,000	7	14.2

Bushler

In Table V data are given on annealed and cold drawn wire of the 55-Nitinol composition. As would be expected, the fibered texture introduced through cold working has the effect of increasing the ultimate tensile strength. Additional cold working markedly increases the strength and simultaneously lowers the tensile elongation (ductility).

Table VI gives very preliminary data on fatigue measured on an R. R. Moore rotating beam test. The resultant "runout" at 25. x 10^6 cycles at 70,000 psi load is unusual in the fact that the load employed represents a high fraction of the expected yield strength of the 56-Nitinol composition bordering on the alloys exhibiting martensitic transformations that this good fatigue life is in part due to the martensitic structural change ($\gamma_1^1 \rightleftharpoons \gamma_2^1$) with each reversal of load. If the martensitic transformations go at the rate of 10^{-4} seconds in this alloy system, then a rotational speed of 5000 revolutions per minute during fatigue testing allow the $\gamma_1^1 \rightleftharpoons \gamma_2^1$ reaction to keep up without difficulty.

Table VII shows three things very clearly about the TiNi-base alloys. First, it indicates that the TiNi compound and alloys based upon this compound are definitely ductile in an unnotched condition. Energy absorption to cause breaking from 38 (60 R_C hardened specimen) to 160 proves beyond question the ductility. Secondly, 55- and 56-Nitinol compositions show little or no indication of becoming brittle when the temperature is decreased to -112°F. This good low temperature impact resistance probably is associated again with the structure change that occurs with decrease in temperature ($\gamma^1 \xrightarrow[\text{temperature}]{\text{lower}} \gamma_1^1 + \gamma_2^1$). Thirdly, these alloys appear to be notch-sensitive, this factor being a design consideration if improvement cannot be made through alloying and treatment.

Environmental Properties

The stoichiometric TiNi composition alloys (55-Nitinol) have been subjected to very preliminary tests to determine their corrosion and oxidation resistance. A summary of these results are given in Table VIII and Figure 12. From the limited data given in Table VIII it can be seen that the corrosion resistance is excellent under normal handling and in marine environment. However, like the commercially available titanium-base alloys, the 55-Nitinol is unsuitable in strong acids and bases at

TABLE V TENSILE PROPERTIES OF COLD DRAWN WIRE

Composition (w/o Ni)	Specimen Condition	Tensile Properties	
		U.T.S.(psi)	Elong.(%)
55.4	Cold Drawn 0.020" dia. wire, annealed at 600°C	177,000	25
55.4	Annealed 0.020" dia. wire, cold drawn to 0.015"	271,000	6

TABLE VI FATIGUE PROPERTIES*

Composition (w/o Ni)	Specimen Cond.	Load(psi)	Cycles	Remarks
56	Hot Wrought	70,000	$25. \times 10^6$	Runout

*Standard R. R. Moore Rotating Beam Test

TABLE V11 IMPACT PROPERTIES (CHARPY)

Composition % Ni	Specimen Cond.	Testing Temp.	Impact (Ft-Pounds)	
			Unnotched	Notched
55	Hot Worked	Room Temp.	117	24
		-112°F	70	17
55.1 (plus up to 0.08 Fe)	Hot Worked	Room Temp.	153	
		-107°F	160	
56	Hot Worked	Room Temp.	105	6
		-112°F	103	6
60	Extruded (Hot)	Room Temp.	56	5
	Hardened (60 R_c)	Room Temp.	38	2

TABLE V111 ENVIRONMENTAL PROPERTIES

OXIDATION

Suitable In Air Up To 600°C

CORROSION

Salt Spray (20% Solution, 95°F, 96 Hours	:	No Effect
Sea Water	:	"
Normal Air Atmosphere	:	"
Normal Handling	:	"
60% H_2SO_4 (150°C)	:	Unsuitable
60% H_3PO_4 (125°C)	:	"
85% KOH (250°C)	:	"
100% NH_3 (100°C)	:	0.0003 Grams Gain (72 Hours)

Buehler

elevated temperatures. Oxidation resistance is shown graphically for a 55-Nitinol alloy in Figure 12, and it is immediately apparent that above about 600°C (1110°F) oxidation occurs at an accelerated rate. At the 800°C and 1000°C temperature, the natural protective coating tends to spall away readily.

Potential Applications

Recognizing the two distinct groups of TiNi alloys and the properties associated with each type, Table IX was prepared. This table should not be considered to cover all of the potential applications in the U. S. Navy, but rather examples of the diversity possible through minor variations of composition and treatment. The properties of being non-magnetic, having lower density and being corrosion resistant are inherent over the entire alloy composition range being considered. Beyond these three universal properties are three additionally listed properties that are composition dependent, and it is on the basis of these latter three properties that Table IX may be divided into two distinct parts.

All of the applications attributed to the composition range from 54.5 to 55.5 w/o nickel are dependent upon the martensitic transformations, where these structural changes promote unusual vibration damping, fatigue life or absorb energy as in the case of armor.

The lower group of applications in Table IX are mainly dependent upon the excess nickel (56 to 62 w/o) over TiNi stoichiometry and its effect in promoting high hardness and strength when required. More is probably known about this group of applications since the U S. Naval Ordnance Laboratory currently has a Bureau of Naval Weapons contract for the development of non-magnetic hand tools from these materials. The 56 to 62-Nitinol alloys, since they should be ideal as tool materials, provide for the first time a suitable non-magnetic material that is capable of being hardened to tool steel hardness by a simple thermal treatment. Preliminary hot forging studies further indicate the desirability of these materials for tools.

Concurrent with this effort, a leading bearing manufacturer is studying (on a proprietary basis) the feasibility of using 60-Nitinol in non-magnetic bearings. Interim progress reports on the bearing investigations indicate almost unqualified success in this application. Further, the authors' concern over suitable dimensional

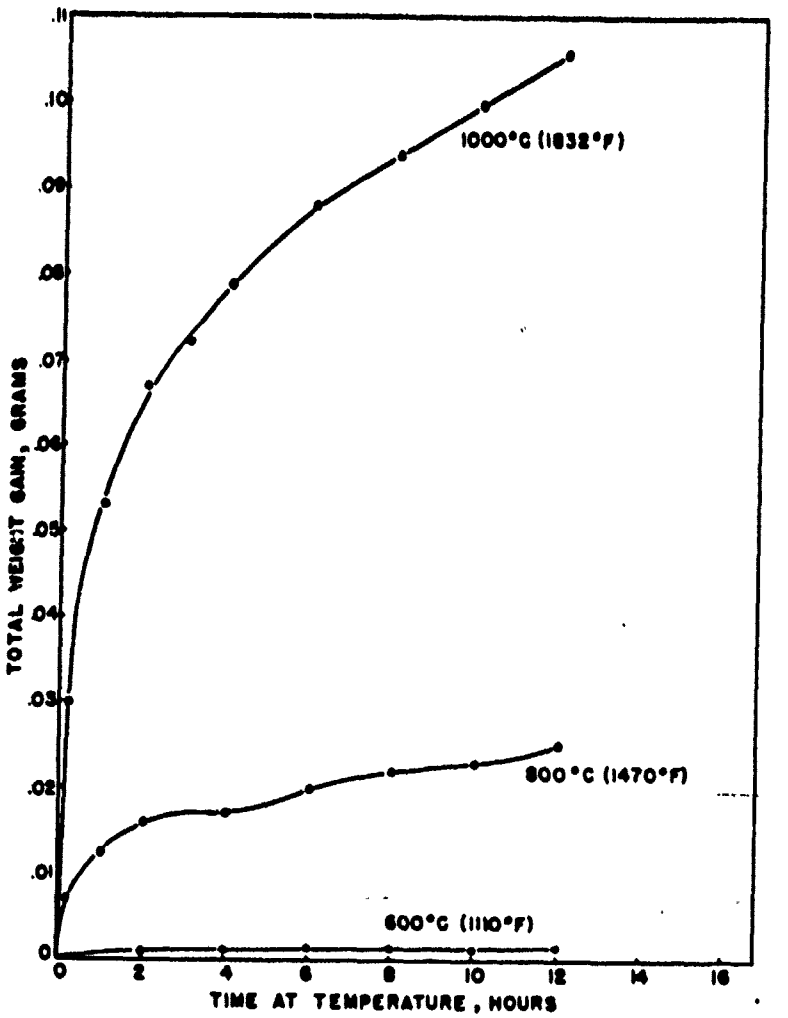


Fig. 12 Curves showing oxidation rates for 55.1-Nitinol at different temperatures. Initial weight of specimens, approximately 6.5 grams.

TABLE 1X POTENTIAL APPLICATIONS

COMPOSITION RANGE (% Ni)	APPLICATION	PROPERTIES						
		Non-Magnetic	Hardening Capability	Lower Density	Corrosion Resistance	Vibration Damping	High Strength	Miscellaneous
54.5 to 55.5	Low Noise Structures			x	x	x	x	
	Temperature Sensing Devices	x			x			$\gamma_1' + \gamma_2' \xrightarrow{\Delta} \gamma'$
	Armor			x	x	x	x	$\gamma' \longrightarrow \gamma_1' + \gamma_2'$
	Hydrofoils			x	x	x	x	
56. to 62.	Non-Magnetic Hand Tools	x	x				x	
	Non-Magnetic Bearings	x	x				x	Abrasion Resistance
	Instrument Components	x	x					Low Elastic Modulus
	Underwater Ordnance	x		x			x	
	Chemical Handling Equipment				x		x	

Buehler

stability for bearing application was very recently dispelled by the following quoted written statement[a]: ". . . The dimensional stability of this material (60-Nitinol) is excellent. We determined dimensional stability as change in boundary dimensions with time. The dimensional stability of parts which have been subjected to mechanical testing appears to be as good as that of parts merely put on the shelf for storage. . ."

Tests performed on actual finished ordnance parts have been gratifying. One such test was performed on spear penetrators for underwater ordnance made as shown in Figure 13. These penetrators measured about four inches in length and had two different penetrating tip configurations. These penetrators were explosively driven through steel plate up to 3/8 inch in thickness, and to illustrate the ruggedness and utility of this material the penetrators were found to be reusable up to as many as five times. Other non-magnetic alloys used (mostly hardened copper-base alloys) were incapable of successfully penetrating the same steel bulkhead even one time without failing structurally.

Summary

In summary, the investigation and development of the TiNi-base compound materials represent an important step forward in metallic materials technology. They provide:

1. A useful engineering material based upon an intermetallic compound.

2. A class of alloys with a very wide and unusual spectrum of properties which have a capability for solving many Navy material needs.

3. A material which is amenable to commercial production.

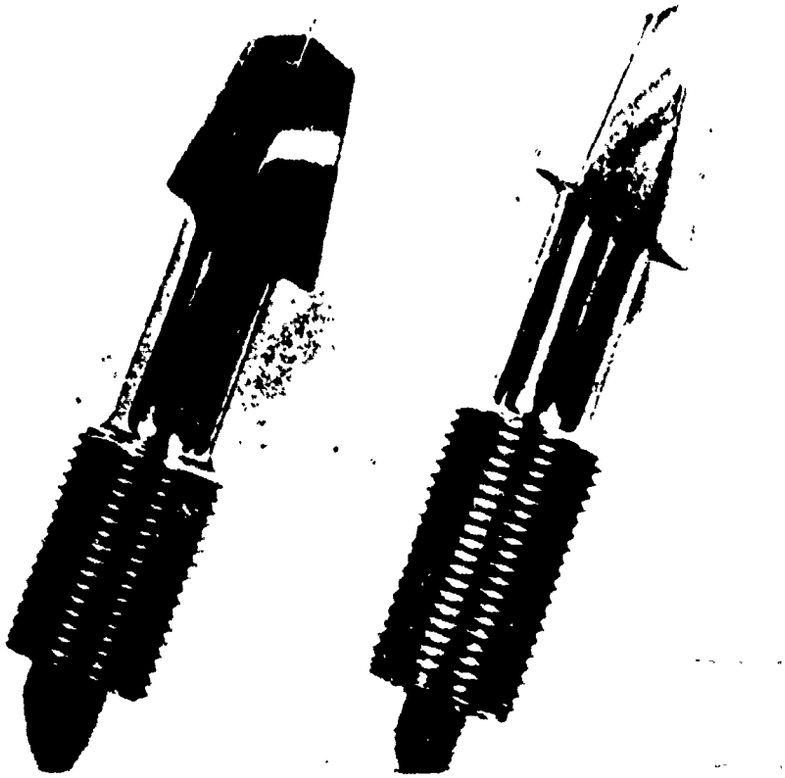


Fig. 13 Two configurations of spear penetrators for underwater ordnance recovery. Penetrators were made of 55.3-Nitinol alloy containing about 0.1 w/o Fe. Actual length of penetrators-about 4 inches.

REFERENCES

1. Metals Handbook, 8th ed, American Society for Metals, 1961.

2. Margolin, H. E. Ence and J P. Nielsen, "Titanium-Nickel Phase Diagram," Trans. AIME, Vol. 197, p. 243, 1953.

3. Duwez, P. and J. L. Taylor, "The Structure of Intermediate Phases in Alloys of Titanium with Iron, Cobalt, and Nickel," Trans. AIME, Vol. 188, p. 1173, 1950.

4. Purdy, G. R. and J. Gordon Parr, "A Study of the Titanium-Nickel System Between Ti_2Ni and TiNi," Trans. AIME, Vol. 221, p. 636, 1961.

5. Hansen, M., Constitution of Binary Alloys, 2nd. ed. McGraw-Hill Book Company, Inc., New York, 1958.

6. Informal communication with C. H. Hannan, Miniature Precision Bearings, Inc. Keene, New Hampshire.

HUMAN QUALITY CONTROL

by

James R. Berkshire
Aviation Psychology Branch
U.S. Naval School of Aviation Medicine

INTRODUCTION

Institutions that modify the behavior of people resemble those that modify materials; and the flow characteristics of a training program, such as aviation training, are enough like those of an industrial production process to permit several useful analogies.

In a typical manufacturing operation, raw materials are selected in accordance with some minimum standards of quality and put through a series of processings which change them progressively, in specified ways, until a desired end product is reached. Usually, at several levels of the sequence, the product is measured to determine whether it is within the tolerances necessary for later stages of the manufacturing process, or for its market. These measurements may be made by gauging, by chemical or metallurgical checks, or by visual or electronic inspections. These checks or inspections may cover all the units or they may be made by sampling; they may be used to monitor the processing, or the intent may instead be that of eliminating early in the total process units which have a high probability of being scrapped later, or of being unsatisfactory to the consumer. What happens in training programs, whether these be civilian (teachers, physicians, scientists, engineers) or military (aviators, submariners, technicians) can be described in very similar language. Candidates for the training must be screened to prevent those with high probabilities of failure from entering training; minimum training standards which will result in the early elimination of potential training and job failures must be imposed at various levels of the training process.

The first step in the accomplishment of these things is the administration of a general intelligence test to all men who apply for naval aviation training. These men have at least two years of college training. Nevertheless, about twelve per cent are rejected for having an intelligence level that would be unbecoming in an officer. The remainder are given a short battery - about two hours - of flight aptitude tests. About 25 to 30 per cent fail these tests. The remainder of these take the flight physical examination--and a further 35 to 40 per cent fail it. Altogether, of every ten applicants, around four pass all the hurdles and qualify for flight training.

All of this, so far, is pretty old hat. Actually there have been no real improvements in selection for aviation training since World War II. The real progress that has been made possible by the computer starts when the men who have been selected for training arrive at Pensacola.

Modern aviation training is expensive--a man who has been selected for training and who leaves the training program after he has completed basic flight training may well represent a lost investment of more than a hundred thousand dollars. While the earliest portion of training (pre-flight ground school) costs only about $100 per week, once the man starts flying the costs mount rapidly, ranging from about $500 per week in a light primary trainer to about $5500 per week in operational types of jet aircraft. Since, under current selection standards, about one-third of the men who start training do not finish, the total amount of money spent on men who fail to complete is quite large. Consequently, procedures which aid in the early identification of men with high failure potentials can save millions of dollars.

In aviation training, as in any training or educational program, whenever a student is having difficulties a decision must be made on whether to drop him or restore him to the program for another chance. Normally, this decision will be made by taking into consideration available information on the past performance of the individual--his grades, his aptitude test scores, his interests, measured or expressed, his judged motivation, etc. In aviation training such measures accumulate rapidly; Table 1 shows the items of quantitative information that become available on non-officer students during their early training. The administrator who must make decisions about marginal students is soon embarrassed by the availability of more performance data than he knows what to do with. In the past administrators have most frequently handled the problem by ignoring most of the available information and relating their judgments to scores on two or three measures in which they had confidence. Quite often this confidence was misplaced.

Table 1

Scores and Grades Available on Non-Officer Aviation Students

1.	Before selection	Aviation Qualification Test
2.	" "	Mechanical Comprehension Test
3.	" "	Spatial Apperception Test
4.	" "	Biographical Inventory
5.	" "	Age
6.	" "	Years Education
7	Second Week	Incoming Mathematics Test
8.	Third Week	Incoming Jump Reach
9.	" "	Incoming Sit-ups
10.	" "	Incoming Speed Agility
11.	" "	Incoming Chins
12.	" "	Incoming Step Test
13.	Sixth Week	Mathematics Final Grade
14.	Ninth Week	Peer Rating
15.	" "	Physics Final Grade
16.	" "	Trampoline
17.	" "	Gymnastics
18.	Eleventh Week	Navigation Final Grade
19.	" "	Study Skills Final Grade
20.	Thirteenth Week	Outgoing Jump Reach
21.	" "	Outgoing Sit-ups
22.	" "	Outgoing Speed Agility
23.	" "	Outgoing Chins
24.	" "	Outgoing Step Test
25.	Fourteenth Week	Engineering Final Grade
26.	Fifteenth Week	Aerodynamics Final Grade
27.	" "	Physiology Final Grade
28.	" "	Naval Orientation Final Grade
29.	" "	Leadership Final Grade
30.	After Pre-Solo Stage	Pre-Solo Flight Grade
31.	After Precision Stage	Precision Flight Grade
32.	After Transition Stage	Transition Flight Grade
33.	After T/28 Precision Stage	T/28 Precision Flight Grade
34.	After Acrobatics Stage	Acrobatics Flight Grade

It seems reasonable to believe that if all of a student's valid past performance measures could be appropriately weighted and combined into a single statement of the probability of his success or failure, the decisions of administrators concerning students might become more accurate. Administrator knowledge of such probabilities should lead to the earlier dropping of men with high probabilities of later failure and the retention of men who have good chances to complete the program. This, in turn, should result in marked improvements in the efficient utilization of training facilities.

PROCEDURE

The punched selection and training records of all men who entered naval aviation training during 1959 were used as basic data. Using a program for computing Pearson correlation coefficients the matrix of the intercorrelations of the selection variables was computed. Then, as each new score, or grade, or rating became available on the men as they moved through training a new matrix containing all available variables was computed on the survivors. Thus we computed a series of fourteen matrices beginning with six variables and ending with thirty-four. Table 1 lists these variables. Using a separate program, bi--serial correlations based on a completed-failed dichotomy were computed for each variable in each matrix. Following this, a program embodying a modification of the Wherry-Doolittle multiple correlation procedure was applied to each successive group of correlation coefficients. This is a procedure that identifies, in their proper order, those variables which combine to give the highest multiple correlation with a criterion--in this case, completion vs. failure. The procedure includes a correction for the additional error resulting from each added variable. It stops adding variables to the prediction formula when the new variable subtracts more error than it adds validity. Following the identification of the variables in each multiple correlation, the next portion of the program computed the weights to be applied to the score on each variable to maximize the accuracy of prediction. Table 2 shows the variables, the numbers of passing and failing cases and the multiple correlation coefficients obtained at each level of training.

For the operational use of these multiple correlation formulae certain simplifications were made. First, wherever two or three successive formulae were identical in variables and similar in weights, only one was used. Second, whenever the last several variables in a multiple added only a few thousandths to the magnitude of the validity coefficient these were dropped from the formula. The variables in these simplified formulae, the periods of training to which they apply and their respective multiple correlation coefficients are shown in Table 3.

Table 2

Variables and Multiple Validities at Successive Stages of Non-Officer Aviation Training

Matrix	Variables	Variables in Multiple	N Completed	N Failed	R
1	1 - 6	1,2,3,4,6	670	243	.37
2	1 - 7	1,2,3,4,6	549	185	.38
3	1 - 12	1,2,3,4,6,8,9,10	549	185	.42
4	1 - 13	1,2,3,4,6,7,8,9,10,13	549	184	.46
5	1 - 17	1,2,3,4,6,7,8,11,13,14,16	513	155	.50
6	1 - 19	2,3,4,6,7,8,1[illegible],13,14,16	513	153	.54
7	1 - 24	no change	513	153	.54
8	1 - 25	3,4,6,8,14,15,16,18,25	513	153	.56
9	1 - 29	no change	513	153	.56
10	1 - 30	3,6,9,10,11,13,14,15,16, 18,25,29,30	502	116	.66
11	1 - 31	2,3,6,8,13,14,15,16,18, 25,30,31	502	83	.57
12	1 - 32	3,6,8,13,14,16,18,25,30, 32	464	64	.59
13	1 - 33	3,6,8,14,16,18,25,30,32, 33	464	55	.55
14	1 - 34	3,8,14,16,18,25,30,31,32, 34	441	44	.58

Table 3

Simplified Multiple Regression Formulae - Non-Officer Students

Stage	Variables in Multiple	R
1st and 2nd weeks	1,2,3,4,6	.38
3rd, 4th and 5th weeks	1,2,3,4,6,8	.42
6th, 7th and 8th weeks	1,2,3,4,6,8,13	.45
9th and 10th weeks	1,2,3,4,6,8,13,14,16	.49
11th, 12th and 13th weeks	2,3,4,6,8,14,16,18	.54
14th, 15th, 16th weeks and Pre-solo prior to hop #9	3,4,6,8,14,15,16,18,25	.56
Pre-solo, hop #9 and after, and T-34 Precision	3,6,13,14,16,18,25,30	.65
Transition	6,8,14,16,18,30,31	.55
T/28 Precision	3,14,16,18,30,32	.56
Acrobatics	3,6,8,14,16,18,30,32,33	.54
After Acrobatics	3,8,14,16,18,30,32,33,34	.57

APPLICATION

In order to transform these multiple validities into probability estimates that could be used to increase the accuracy of decisions about individual students the following steps were taken:

1. For each student who had entered training during 1959 a regression score was computed for each of the eleven stages of training shown in Table 3. Thus, at the first stage (1st and 2nd weeks), 1078 regression scores were computed by multiplying each student's scores by the appropriate weights and summing the products. This was repeated for each later stage for those students still in the program.

2. At each stage the frequency distribution of the regression scores of men who subsequently graduated from training was compared with the frequency distribution of the scores of those men who subsequently failed to complete training. By dividing these distributions into five or six segments, and by determining the relative numbers of completing and non-completing students within each segment it was possible to establish empirical probabilities of success for each regression score.

3. The results were put into tables of variables, beta weights, regression scores and probabilities of success for use at each level of training (see Figures 1 and 2).

The instructions that have been issued to naval aviation training personnel concerning the use of these materials are as follows:

"This instruction sets forth a procedure by means of which the computer--determined probability of success of any student at any point in training can be ascertained.

Step 1 - When a student is in trouble and a decision must be made as to whether he should be dropped or given further training time, call Mr inside 5137 and ask for 'Student Prediction.' Give the student's name, pre-flight class, and his point in training. Give your telephone number and name.

Step 2 - Within approximately 15 minutes they will call you back and give you the computed probability that this student will succeed (or fail) if you restore him to training. This report will be something like 'Based on his marks so far, four out of five such students fail to complete training' or 'He falls into a group of students in which three out of four succeed.'

Variables	Mean	S.D.	Weight
Spatial Appreception	20.60	5.48	.27
Biographical Inventory	33.20	8.20	.10
Education	6.74	1.22	.41
Peer Rating	50.30	10.86	.20
Navigation	48.92	6.80	.34
Trampoline	30.15	5.98	.17
Jump Reach (1)	11.03	2.77	.20
Engineering	49.97	8.25	.24
Physics	48.13	9.63	-.10

Regression - Probability Table

Score	Per Cent	Will Complete	Won't Complete
60.0 or more	9	7	1
53.0 - 59.9	43	3	1
49.0 - 52.9	28	1 1/2	1
45.0 - 48.9	14	1	1 1/4
44.9 or less	6	1	4 1/2

N = 967 R = .56

No completions below 40.9 - ten failures

Fig.1. Tables for Determining Probability of Graduation - Non-Officers 14th, 15th and 16th Weeks

Berkshire

Variables	Mean	S.D.	Weight
Spatial Appreception	20.73	5.45	.21
Education	6.73	1.22	1.20
Math Final	48.33	9.66	.22
Peer Rating	50.50	10.37	.16
Navigation	49.05	6.68	.17
Engineering	50.41	8.07	.15
Trampoline	30.31	5.98	.15
Pre-Solo	294.71	11.67	.44

Regression - Probability Table

Score	Per Cent	Will Complete	Won't Complete
194.0 or more	8	11	1
186.0 - 193.9	29	6	1
179.0 - 185.9	36	3	1
174.0 - 178.9	19	1	1
170.0 - 173.9	5	1	2
169.9 or less	3	1	24

N = 867 R = .65

No successess below 169.0 - eighteen failures

Fig.2. Tables for Determining Probability of Graduation - Non-Officers Pre-Solo, Hop #9 or after, and Precision

Step 3 - These reported probabilities are based upon all the significant grades and ratings, weighted in accordance with their relative importance to later success. When you use this probability in making a decision about the man do not give additional consideration to any individual scores, grades, or ratings. They are either already in the probability figure or they are not valid predictors of training success. On the other hand, it is not intended that these probabilities be the sole basis for decision. Administrative officers and boards are still expected to consider each case on its unique merits, recognizing that all of the important aspects of a case or of a man may not be reflected in this probability statement."

For the past two years the average length of time that students who failed to graduate remained in the training program was around thirty-six weeks. It is expected that the procedure that I have described will reduce this average time by the earlier identification of men with high probabilities of later failure. Given our present training load and percentage of non-graduates, each week that we can reduce this average of thirty-six weeks will result in a savings in training costs of around two million dollars.

PILOT PERFORMANCE DURING CENTRIFUGE SIMULATIONS OF ACCELERATION ENVIRONMENTS

Randall M. Chambers and W.S. Wray, CAPT, MC, USN
Aviation Medical Acceleration Laboratory
U.S. Naval Air Development Center
Johnsville, Pennsylvania

To date, a major research effort has been directed toward the definition and solution of physiological problems associated with various acceleration environments. Many of these investigations have been excellently reviewed by Gauer and Zuidema (1), Price (2), and Wood (3). A lesser amount of research has been directed toward human performance problems encountered in acceleration environments, although papers by Brown (4, 5), Chambers (6, 7, 8), Chambers and Doerfel (9), Chambers, et al (10), Chambers and Nelson (11), and Chambers and Hitchcock (12) have reviewed some of these problems. The majority of these problems pertain to the capabilities of pilots for maintaining control of their vehicles during normal launch and reentry missions, and during certain types of emergency escape conditions. The present paper reports the results of some centrifuge simulation studies of manned space flights and proposed spacecraft. The studies were conducted on the Aviation Medical Acceleration Laboratory (AMAL) Human Centrifuge in support of USN, NASA and USAF space projects, and acknowledgment is made to the contributions of the many participating pilots, scientists, and engineers.

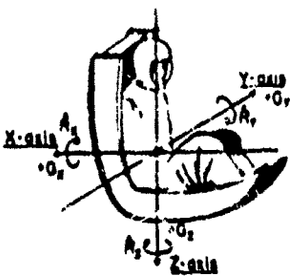
ACCELERATION ENVIRONMENTS

Most acceleration environments may be described in terms of magnitude, direction, rate of onset and decline, pattern complexity (composition of various linear and angular components), and duration. The severity of any particular acceleration environment is dependent upon many physical, physiological, and psychological variables, among the most important of which are the physical characteristics of the accelerations themselves.

Since a consistent nomenclature has not been adopted for universal use by engineering, biological and behavioral science groups (Clark, et al, 13), it is important to describe briefly the nomenclature system used in this paper. This system is shown in Figure 1. It is based on the displacement of body fluids and the heart with respect to the skeleton when the pilot is positioned approximately as shown with respect to his X, Y, and Z axes. G is used in this nomenclature as a unit of reactive force (Gauer and Zuidema, 1), and consequently this terminology emphasizes body reactions to the accelerations rather than the accelerations themselves. The positive and negative signs along each of the three axes are directions of reactive displacements within the human body — $+G_x$ means that the heart is displaced from the chest toward the back with respect to the skeleton, and $-G_x$ means that the heart is displaced toward the chest, sometimes referred to as "eyeballs in" and "eyeballs out" accelerations, respectively. The positive physiological Y axis is right to left, and consequently, $+G_y$ means that the heart moves toward the left with respect to the skeleton. If the heart moves to the right with respect to the skeleton, this is called $-G_y$. These two types of accelerations are sometimes called "eyeballs left"and "eyeballs right", respectively. The Z axis extends along the spinal column — $+G_z$ means that the heart is displaced caudally along the spine, and $-G_z$ means that the heart is displaced toward the head. These accelerations are sometimes called "eyeballs down" and "eyeballs up".

Similarly, the angular acceleration moving around the X-axis, Y-axis, and Z-axis are also described in this diagram. Thus acceleration about the X-axis means that the acceleration forces are causing the heart to roll left in the chest. This is designated as $\dot{R}_x$. Acceleration about the Y-axis means that the acceleration forces are causing the heart to pitch down within the chest. This is designated as $\dot{R}_y$. Acceleration about the Z-axis means that the acceleration forces are causing the heart to yaw left in a direction about the Z-axis, and this is designated as $\dot{R}_z$. The negative sign is conveniently used to indicate conditions in which the acceleration forces are causing the heart to be displaced in the opposite directions around the X-axis and Z-axis.

This anatomical nomenclature permits a physiological description of the forces acting on man with respect to his own X, Y, and Z body axes. Most other nomenclatures pertain to the forces acting on the X, Y, and Z axes of the vehicle, and consequently a description of the man's specific environment within the vehicle is difficult, especially if he changes body positions during flight.



(Body Fluids, Heart Displacement, With Respect to Skeleton)
Linear Acceleration Modes

Description of Heart Motion

Actual	Other Descriptions			Symbol	Unit
Towards spine	Eye-balls-in	Chest-to-back	Backward facing	$+G_X$	g
Towards sternum	Eye-balls-out	Back-to-chest	Forward facing	$-G_X$	g
Towards feet	Eye-balls-down	Head-to-foot	Headward	$+G_Z$	g
Towards head	Eye-balls-up	Foot-to-head	Footward	$-G_Z$	g
Towards left	Eye-balls-left	——	Rightward	$+G_Y$	g
Towards right	Eye-balls-right	——	Leftward	$-G_Y$	g

$$NG = \frac{a}{g} = N_1 G_X + N_2 G_Y + N_3 G_Z$$

$$N^2 = N_1^2 + N_2^2 + N_3^2$$

Angular Acceleration Modes

Acceleration about X-axis (roll axis)	$\dot{R}_X$	rad/sec^2
Acceleration about Y-axis (pitch axis)	$\dot{R}_Y$	rad/sec^2
Acceleration about Z-axis (yaw axis)	$\dot{R}_Z$	rad/sec^2

(Angular acceleration is positive or negative by right hand rule)

Fig. 1 Physiological displacement acceleration nomenclature

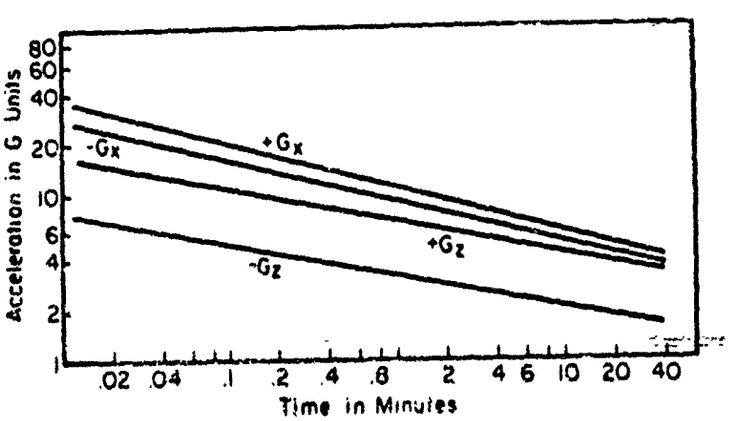


Fig. 2 Average acceleration tolerances for transverse supine acceleration ($+G_X$), transverse prone acceleration ($-G_X$), positive acceleration ($+G_Z$), and negative acceleration ($-G_Z$)

In acceleration physiology, it is convenient to measure acceleration $\underline{a}$ as a multiple of $\underline{g}$, the standard unit of acceleration due to gravity, and any force $\underline{F}$ as a multiple of the standard weight $\underline{W}$ of the body upon which $\underline{F}$ is acting. However, because of the physiological importance of the position of the human body with respect to the acceleration force, the directional aspects of the acceleration quantity are considered differently from those used in physics. In acceleration physiology, only the magnitudes of the vectors $\underline{g}$ and $\underline{W}$ are employed. The directions of the vectors $\underline{a}$ and $\underline{F}$ are independent of the fixed directions of the physical quantities $\underline{g}$ and $\underline{W}$. It is desirable to employ a single symbol to represent the common ratio formed by normalizing $\underline{F}$ and $\underline{a}$ with respect to the standard gravitational values. Thus, G is defined as a ratio of forces, or, as a ratio of accelerations (Dixon and Patterson, 14) as follows:

$$G = \frac{a}{g}, \text{ or } G = \frac{F}{W}$$

More discussions and comparisons of various nomenclatures are described by Clark, et al (13), Gauer and Zuidema (1), Gell (15) and Chambers (7).

PHYSIOLOGICAL TOLERANCE TO ACCELERATION STRESS

Physiological tolerance is one of the most important concepts in acceleration research. It is usually defined as the physiological ability to sustain, endure or withstand the acceleration stress. The tolerance criterion is usually somewhat arbitrary, and it encompasses both physiological and psychological functioning. The concept includes the usual physiological response brought about by the intensity and duration characteristics of a stimulus, such that the point of response may be separated from the point of no response by a simple stimulus strength-duration curve. Many different kinds of criteria have been used to measure G tolerance. These include EKG abnormalities in rate or wave form, cardiovascular response survival time, chest pain, grayout, blackout, and unconsciousness. There are excellent reviews of the physiological tolerance problem in the scientific literature, and the reader is referred to these for more exhaustive considerations of the tolerance problem (Eiband, 16; Webb, 17; Gauer and Zuidema, 1).

Figure 2 presents a set of curves which we have extrapolated from approximately 20 experiments. The figure shows some of the most important relationships for magnitude of acceleration and duration time for positive acceleration ($+G_z$), negative acceleration

($-G_z$), transverse supine acceleration ($+G_x$), and transverse prone acceleration ($-G_x$). The plot is partially theoretical since data along some of the intermediate points are not available. The figure represents averages, and not necessarily maximum tolerance levels. The figure shows that the acceleration load which a subject can sustain for any given duration time is higher for transverse supine ($+G_x$) acceleration than for the remainder of acceleration vectors. The primary limiting factor for positive transverse acceleration is respiration difficulties and fatigue. Transverse prone acceleration ($-G_x$) is next, and the primary limiting factors are visual decrement and periorbital pain. Positive longitudinal acceleration ($+G_z$) can be endured at lower level, and the data in this figure assume the presence of a G-suit to assist in maintaining blood flow in the trunk and head areas. Visual decrement (visual grayout and blackout) without necessarily the presence of pain, are the primary limiting factors for this acceleration load. Negative longitudinal acceleration ($-G_z$) can be sustained for only short periods of time, and for only relatively low acceleration levels. Excessive pain in the head and eyes and possible severe cardiac damage may result from negative longitudinal acceleration.

Figure 3 summarizes the results of an experiment in which a group of healthy pilots attempted to sustain relatively high $+G_x$, $-G_x$, and $+G_z$ acceleration vectors for long periods of time using the Ames restraint system, a portable, adjustable type designed for all three vectors (Smedal, et al, 18). The system was designed to provide the pilot with not only physiological G protection, but also with minimal restriction for performing operations, providing restraints for the arms, feet, and head so that these body members could be maintained secure, yet relatively free to move, during exposure to G. The criteria for these runs were: (a) medical, e.g., the medical officer decided that the runs should be terminated because of cardiac or respiratory problems, or (b) the pilot's tracking performance became unsatisfactory, as judged by a performance monitor. The figure presents some unusually outstanding time-tolerance to acceleration centrifuge runs along three different acceleration vectors. This particular figure represents results of prolonged high-G acceleration tests in which experienced pilots served as subjects. They continuously performed piloting tasks in acceleration fields during steady-state G exposures in which performance proficiency as well as physiological tolerance were used as tolerance criteria.

PERFORMANCE TOLERANCE TO ACCELERATION STRESS

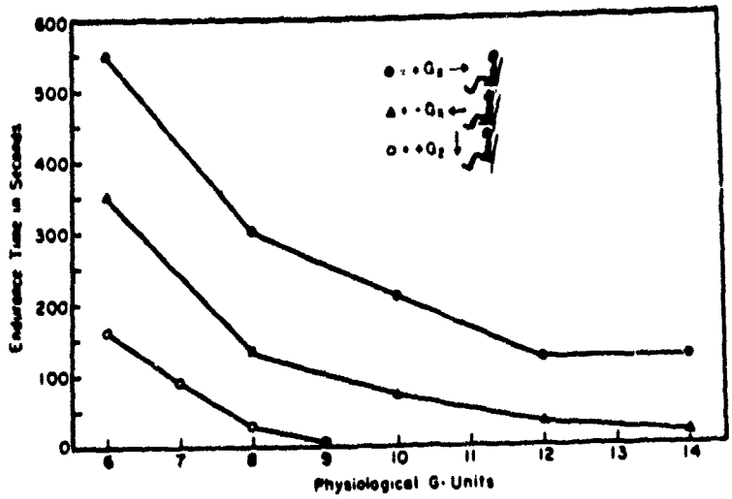


Fig. 3 Endurance time in seconds versus physiological G units for a group of highly motivated pilots

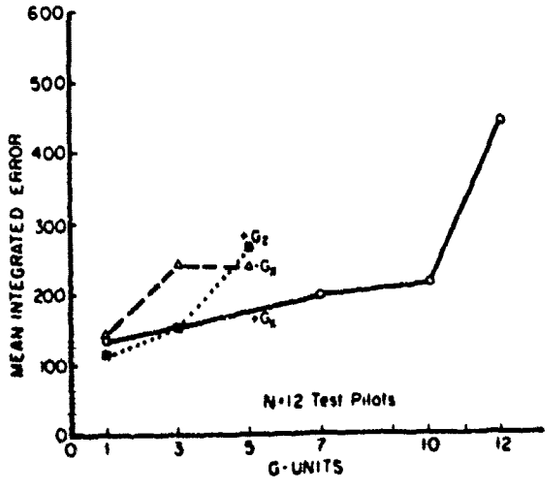


Fig. 4 Mean Integrated Error for eye-balls in ($+G_x$), eye-balls out ($-G_x$) and eye-balls down ($+G_z$) accelerations for 12 test pilots during sustained tests on the human centrifuge

In addition to the physiological tolerance limits which define the end points for reliable functioning for any particular physiological system during exposure to acceleration stress, there are also performance tolerance limits which define the end points for reliable functioning of any particular overt behavior system during acceleration (Chambers and Nelson, 11). The physiological and performance tolerance limits may be functionally related, but they are not necessarily the same. Performance tolerance limits usually indicate the G amplitude level or time during which a pilot may satisfactorily perform a given task (see Figure 4). The specification and development of performance tolerance maps which show impairment as a function of physiological acceleration stress are dependent upon the identification and quantification of performance errors so that the amount of impairment of the particular human ability in question may be indicated.

It has long been known that acceleration may produce subtle confusion, disorientation, memory lapses, and some loss of control of voluntary movements. However, there is a need for meaningful and quantifiable expressions of accuracy, speed, and consistency of piloting performance during exposure to these adverse environments.

There is not time during this conference to present a detailed account of the amplitude, frequency, and duration parameters of error performance as we have observed them during acceleration studies. However, there is time to present a list of eleven different types of error performance which test pilots and astronauts frequently make during stressful portions of acceleration runs, but which they do not generally make during static control runs. These are not specific to any particular kind of piloting tasks. Rather, it is believed that these eleven types of errors are common to piloting behavior under conditions of intense or prolonged acceleration stress. They are meaningful expressions of accuracy, speed, and consistency of behavior.

(a) increase in error amplitude

(b) lapses, or increasing unevenness, and irregularity, in performing the task

(c) performance oscillations

(d) falling off in proficiency, or failure to respond, in some part of the task, and not in other parts of the task, while maintaining proficiency in other parts

(e) changes in timing or phasing of task components, as may occur for either a multidimensional task or in a sequential task

(f) inadvertent control inputs

(g) changes in the rate (or frequency) of performance, including response lags

(h) initiation of performance nonessential to the task

(i) over-controlling and under-controlling, as during a transition phase

(j) failure to detect, or inability to sense, perceive, and/or otherwise retrieve stimuli and other information

(k) errors in retrieving, integrating, storing, processing or reporting information

G-PROTECTION

In addition to the concepts of physiological and performance G tolerance, an important concept is that of G-protection. There are many kinds of G protection, including form-fitted contour couches, net couches, G-suits, water suits, a large variety of straps, restraints, bindings, and foams. G-tolerance may be raised considerably, depending on the nature of the protective device used. G-protection equipment also influences the kinds of tasks which a person may perform during acceleration exposure since, in addition to providing support so that the subject may withstand the acceleration and consequently perform better, it also restrains and restricts the subject's movements and the kinds of operations he may perform. G-protection may either facilitate or impair performance, depending on how it is designed with respect to the pilot and his acceleration environment.

The G-protection system which is receiving most intensive study at the present time is the contour couch. During the past four years, AMAL working closely with NASA, has been instrumental in developing and testing a variety of contour couches which could be used for operational use. A family of these couches is shown in Figure 5. Each couch is individually molded for a specific person. The one at the far left was developed in cooperation with the NASA High Speed Flight Center at Edwards Air Force Base, California, and was found to be satisfactory for acceleration loads extending to +15 G_x during which time the pilot was required to fly complex two-stage boost orbital missions. The second couch was a design developed for the Mercury Astronauts and this particular couch for Astronaut Carpenter, permitted acceleration runs to +14 G_x, and some of the astronauts who were fitted to these couches achieved runs to 19 G_x. These couches were developed in cooperation with NASA Langley Research Center.

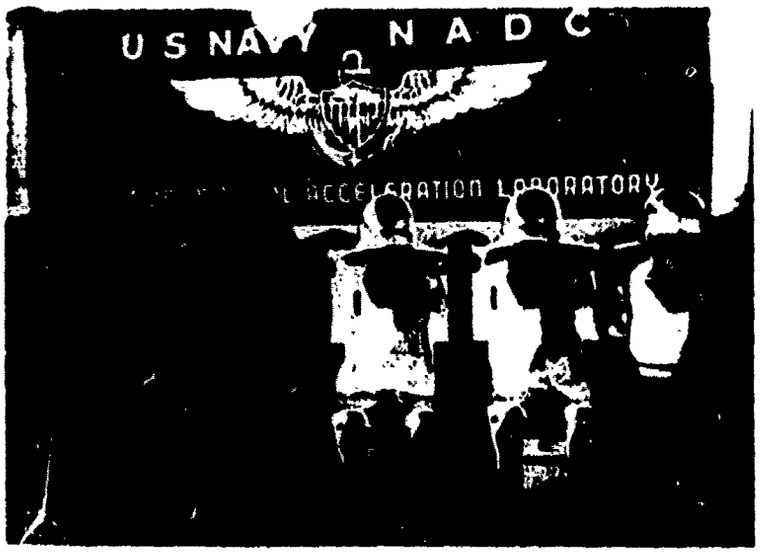


Fig. 5 Examples of individually molded contour couches used in pilot performance studies

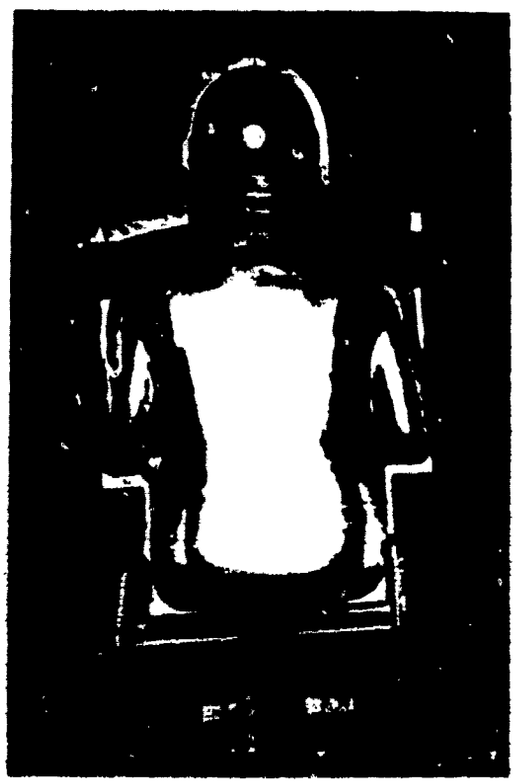


Fig. 6 Modified contour couch used for astronaut training

In the center is a couch developed in cooperation with the NASA Ames Research Center. The primary feature of this couch was that the feet could be freed so as to allow the pilot to use toe pedals. Pilots have successfully used couches of this type to as high as +14 G_x without losing control of relatively complex piloting tasks. The fourth couch represents a model which was developed for permitting the use of the Mercury full-pressure suits. This model of couch was used in most of the centrifuge acceleration training projects for the Mercury Astronauts. These couches were developed in cooperation with NASA Langley Research Center and the McDonnell Company. The last couch, represents the final design used in some of the centrifuge training programs for the Mercury Astronauts, showing slight additions of an inner-liner and slightly modified head support. This couch design was used in the early Mercury flights and has been found effective for tolerating acceleration loads up to 14 G_x without losing control of a relatively complex reentry task.

Figure 6 shows the contour couch used to provide acceleration refresher training for Astronauts Walter Schirra and Gordon Cooper. Note that the original leg restraints have been removed and a relatively light lateral support has been substituted. This allows the astronaut more freedom of leg movement during prolonged periods of confinement while in his capsule. This particular seat has also been used on a number of basic research studies in which it has been found that the addition of positive pressure breathing and 100% oxygen slightly increases acceleration tolerance time to transverse acceleration.

One of the basic needs in the area of G-protection is a "universal fit" contour couch. In Figure 7 a contour couch is present which has been tested on test pilots from the USAF Aerospace Research Pilots' School at Edwards Air Force Base, California, and on other volunteer subjects ranging from heights of 5'2" to 6'3", and in weight from 110 pounds to 219 pounds. The couch consists of three layers of foam of different masses, arranged so as to provide maximum G-protection and comfort at all G levels. Note its deep lateral supports and the smooth head supports. This couch has been tested for both lateral and transverse acceleration up to 15 G. It has been tested on a wide variety of subjects and found to be extremely comfortable for moderately high transverse acceleration loads.

A variety of unusual restraints have been used in acceleration studies including net couches, water-tanks, water-capsules,



Fig. 7 AMAL universal G couch designed to provide $+G_x$ protection for all centrifuge subjects

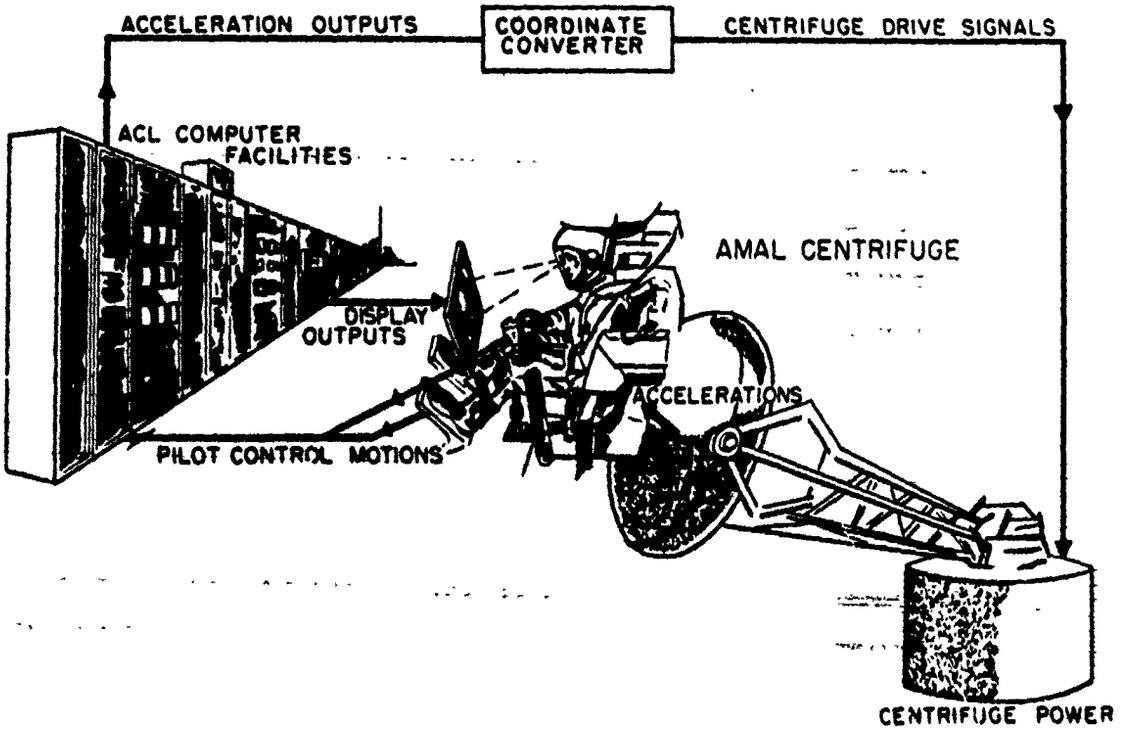


Fig. 8 Diagram of the human centrifuge and associated computer facilities to provide closed-loop flight simulation of space vehicle environments

and procedures for completely foaming the human subject. These are reviewed in articles by Bondurant, et al (19), Clark and Gray (20), Weaver, et al (21) and Leverett, et al (22).

DYNAMIC FLIGHT SIMULATION

Many of the physiological and performance problems expected to be encountered in flight may be studied by means of simulation techniques. By using centrifuges, rocket tracks, and other acceleration devices, it is possible to produce some of the acceleration conditions of real flight. Unconstrained motion with aircraft and spacecraft involves six degrees of freedom which may be conveniently expressed in terms of six components, three of which are linear accelerations and three of which are orthogonal angular accelerations. For any given aircraft or spacecraft, some of these components are more important than others, and the ways in which these are combined determine the complexity of the pilot's acceleration environment. It is possible to express these linear and angular accelerations in terms of the amplitude, direction, rates of onset and decline, duration and pattern of acceleration components, using the acceleration nomenclature which has been described already. This permits a comprehensive description of the acceleration environment with respect to the human pilot, and many aspects of the acceleration environment can be simulated on the human centrifuge.

A schematic diagram of the AMAL Human Centrifuge and some of its associated computer equipment is shown in Figure 8. When the responses of the pilot are included within the driving mechanism of the acceleration device so that the accelerations he receives from moment to moment vary as a function of his behavior, an interesting type of interaction effect occurs, since the pilot's behavior also varies as a function of the acceleration he experiences. A photograph of this device, including a Mercury Astronaut in his pressure suit, is shown in Figure 9. After the astronaut enters the centrifuge, the hatch is closed, and the atmospheric pressure may be regulated in order to simulate some of the environmental conditions which may be encountered during normal and emergency flight.

An astronaut, in his cockpit within the centrifuge gondola, is shown in Figure 10. The pilot within the gondola of the centrifuge is provided with an instrument display panel, a control device, and other piloting equipment, as may be required. The pilot operates his control devices in response to information presented on the instrument panel and cues which he receives during the acceleration.



Fig. 9 Pilot entering the two-gimbaled gondola at the end of the 50-foot arm of the AMAL human centrifuge



Fig. 10 Project Mercury astronaut performing a capsule attitude control maneuver during a simulated space flight.

The analog computers are used to close the loop between the pilot, his displays, his controls, and the centrifuge accelerations. Thus, the control movements which the pilot makes are converted into electrical signals and fed into the analog computer, which continuously generates the flight problem and provides solutions which result in output signals. Some of the signal outputs are transformed by a coordinate conversion system into appropriate centrifuge control signals which regulate the power voltages to the arm and gimbal system of the gondola. Simultaneously, other signal outputs are fed to the pilot's instruments.

The pilot-centrifuge-computer system described above consists basically of two closed-loop systems: one connecting the pilot's control responses with the driving system of the centrifuge, and the other connecting the pilot's control responses with the driving mechanisms of the indicators on the pilot's instrument panel. (Hardy, et al, 23; Chambers and Doerfel, 9; Chambers, 7.)

This procedure has been used in a number of projects, such as the X-15, Mercury, a number of basic research studies, and X-20. The cockpit installation used in X-20 on the centrifuge is shown in Figure 11.

During a typical simulation program on the AMAL centrifuge, there are from 3 to 9 duty stations at which various types of recordings are taken. These recordings include psychological, performance, medical and engineering data. Sometimes, a large analog computer system records performance error as a function of the programmed task and may, if desired, convert the analog scores to integrated error scores or to digital readouts on IBM cards. Another computer system computes means and variability for each run as the run proceeds, thereby providing detailed in-line scoring as each run progresses. In addition, there are several additional data processing systems available for special purpose analysis such as a 14-channel magnetic tape recorder. For example, Figure 12 shows a performance monitoring system. At this station, in-line data recording and data processing is provided by feeding the responses through a small analog computer system which simultaneously yields individual means and standard deviations of the subject's performance on several task components. Figure 13 shows a block diagram of some of the performance analyses which were done in support of the recent X-20 project.

If programmed appropriately, the human centrifuge may be used as a dynamic simulation device in which physiology and



Fig. 11 Dyna-Soar pilot inserted into gondola reviewing mission profile



Fig. 12 Performance monitoring station

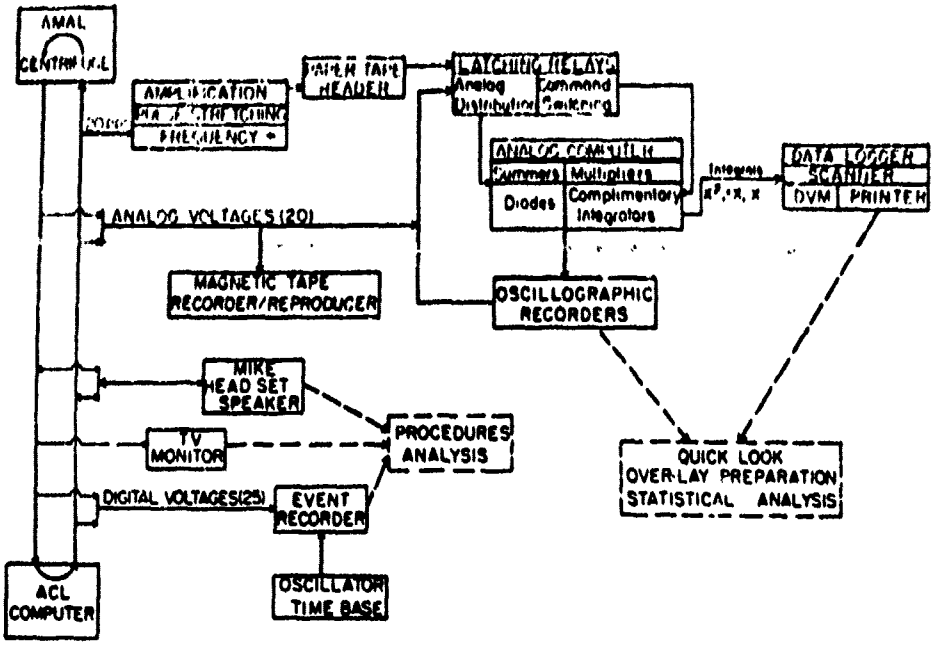


Fig. 13 Diagram of performance monitoring station operations for Project Dyna-Soar

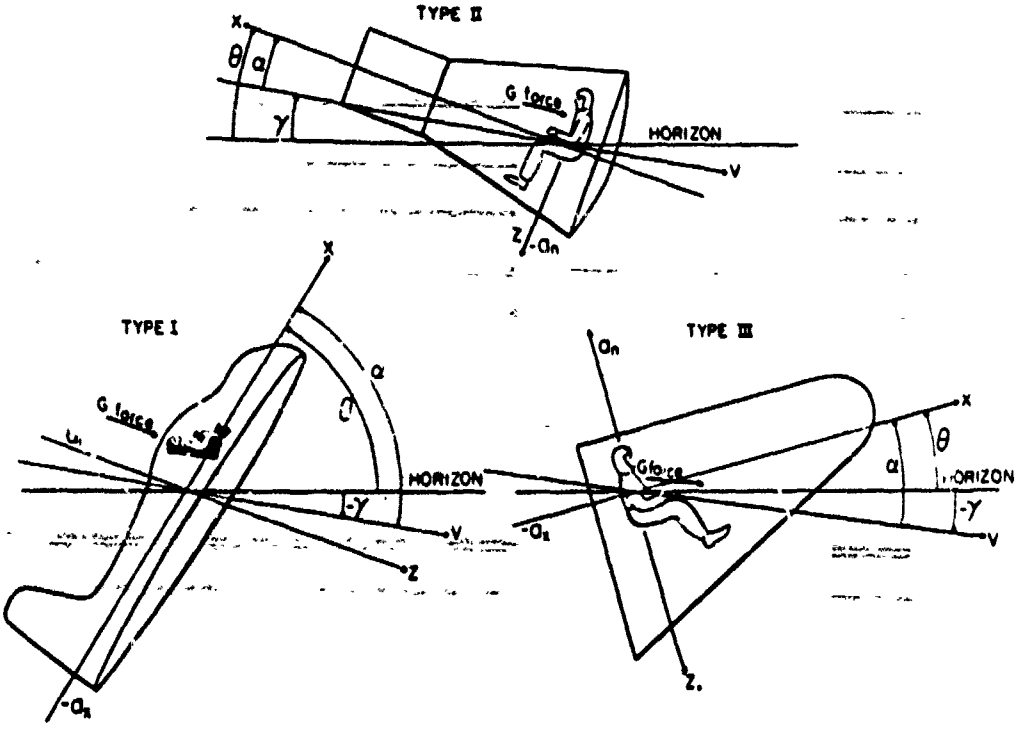


Fig. 14 Three basic types of proposed space vehicle configurations within which human factor problems have been studied at the AMAL

performance of pilots, and the behavior and effectiveness of cockpit instruments, may be studied and evaluated. If programmed to simulate specific types of aerospace vehicles during definite portions of flight maneuvers, the human centrifuge may serve as a very useful tool for identifying and investigating some of the human factor problems associated with a wide variety of the acceleration aspects of flight. The effects of acceleration on pilot physiology, pilot performance and pilot ability to use specific controls, displays and escape equipment may be investigated. In addition, if the centrifuge is instrumented with appropriate environmental conditions such as atmospheric pressure, pressure suit, oxygen and other gaseous conditions, and computer control of the behavior of both centrifuge and the panel instrument, the centrifuge serves as a very useful tool for studying the effects of combinations of conditions which a pilot may expect to encounter during any given particular acceleration phase of his flight. Consequently, the centrifuge may also serve as a very effective tool for studying the integrated performance of the pilot and many selected aspects of cockpit, displays, controls, and environmental conditions. The interaction effects may serve as convenient indications of complete man-machine systems performance. Finally, the human centrifuge is an extremely useful training device for acceleration aspects of complex flight missions.

To date, most of the simulation programs have been concerned with human factors problems relating to each of three basic types of space vehicles, as illustrated in Figure 14. At the bottom left of the figure is shown a high drag, variable lift winged vehicle (Type I). During reentry in this vehicle, the pilot performs while sitting in a normal flying position and receives positive acceleration forces along his Z axis. At the top of the figure is shown the high drag capsule (Type II), in which the pilot performs while sitting facing backwards with respect to his velocity, and in which he receives supine transverse acceleration forces primarily along his X axis. To the right is shown a sketch of a glide capsule (Type III) in which the pilot performs while facing forward in his vehicle, and in which he receives G forces which are resultants of $-G_x$ and $+G_z$ and $-G_x$ components. During reentry, the pilot in the high-drag, variable lift winged vehicle is pressed down into his seat; the pilot in the high drag capsule is pressed against the back of his seat; and in the glide vehicle, the pilot is pressed against his shoulder straps. At our laboratory, simulations in the first vehicle type have been conducted to acceleration levels as high as $+10\ G_z$; simulations in the second vehicle have gone to as high as $+21\ G_x$; and simulations in the third vehicle have gone as high as $-14\ G_x$.

Studies involving tolerance times to steady-state accelerations as programmed on the centrifuge have shown approximate performance time tolerances for pilots in each type of configuration. These are shown in Figure 15 which illustrates approximate tolerable limits only for the times and gravities indicated. These approximations are based on data obtained at AMAL during centrifuge simulations of proposed space vehicles. The tolerance times in heavy print are those resulting when the pilots were protected in advanced restraint systems, which included individually molded contour couches and associated chest, arm, leg, and head restraints.

ASTRONAUT TRAINING

The AMAL Human Centrifuge has been found to be a very useful device for astronaut training. Since 1958 it has been one of the major training devices for preparing the Mercury Astronauts for the acceleration phases of their suborbital and orbital space flights. Figure 16 summarizes some of this training effort. The active Mercury-type instrument panel, Mercury-type side-arm controller, complete environmental control system, and remotely-controlled centrifuge drive system permitted extensive training on a wide variety of piloting tasks and emergency conditions during exposure to the various acceleration profiles for Redstone, Atlas, and abort maneuvers. The instrument panel, shown in Figure 17, was used in a variety of ways to acquaint the astronaut with instrument malfunctioning, visual perception changes during acceleration, flight problems, and the operation of his telelights, knobs, and handles. The association of telepanel indicator lights with acceleration levels and capsule events constituted a major training effort and, as may be seen in Figure 18, the acceleration profile and its associated events constituted a complex sequence of events.

A complete environmental control system, pressure-suit, contour couch restraint system, 100% oxygen, and biosensors were provided. When the gondola was sealed and depressurized to 5 pounds per square inch, the interactions with many different spaceflight conditions were experienced by the astronauts. During the course of five training programs at this facility, the astronauts received practice in straining in order to maintain good vision and physiological functioning under high G loads, and in developing breathing and speaking techniques during high G launch, reentry, and abort stress. Experience in tumbling and oscillations during relatively high G exposures was also provided. The astronauts were given extensive practice in controlling their simulated vehicle during reentry and other phases of their simulated flights. They

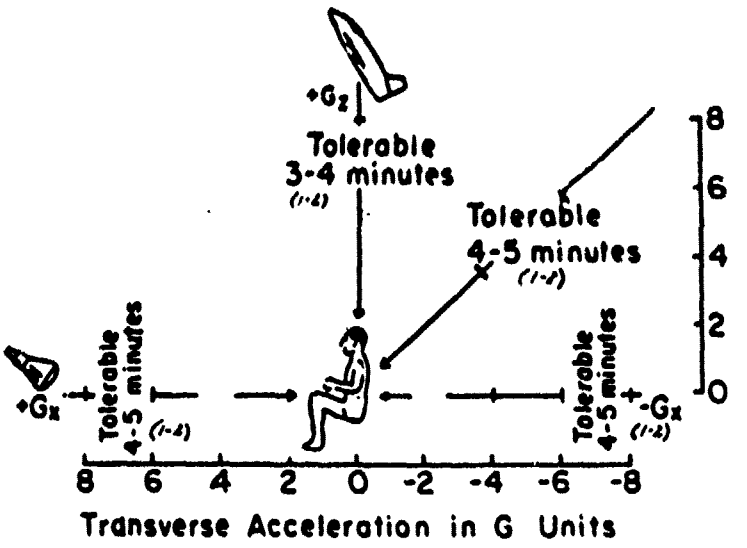


Fig. 15 Approximate performance time tolerances for experienced pilots performing piloting tasks in acceleration fields during centrifuge re-entry simulations of space vehicles

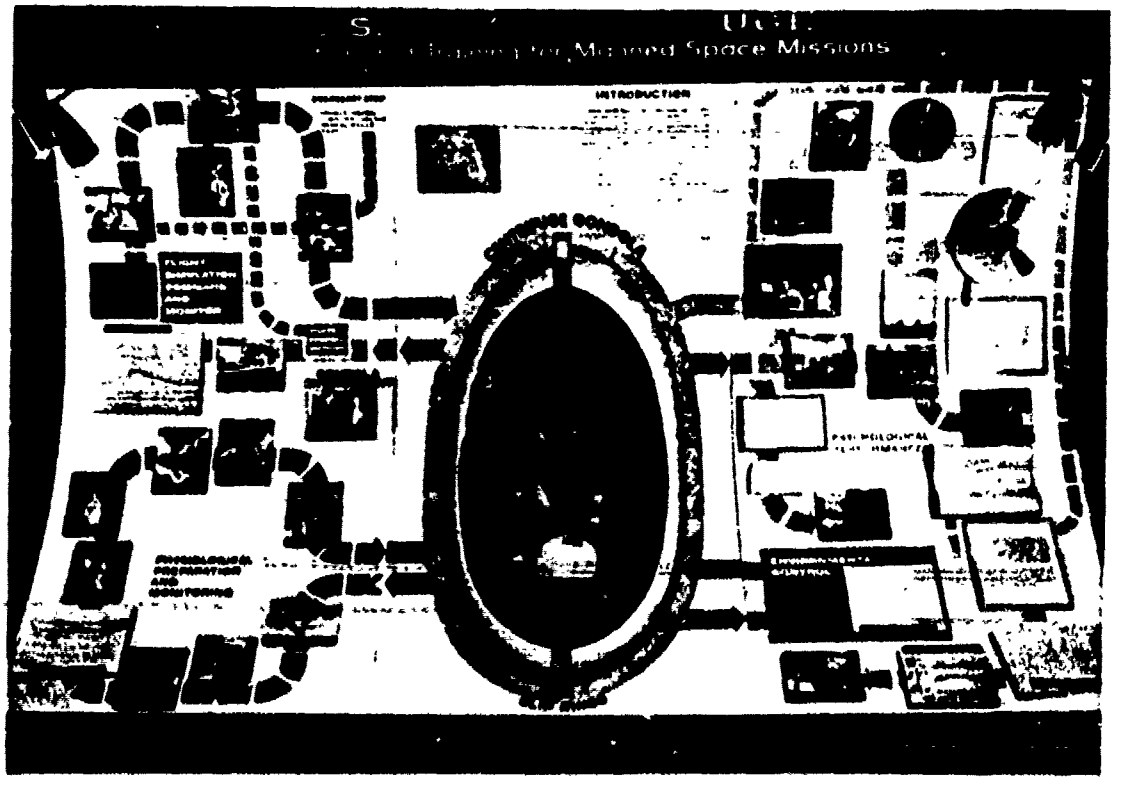


Fig. 16 Model of U.S. Navy Centrifuge - Simulation and Training for Manned Space Missions

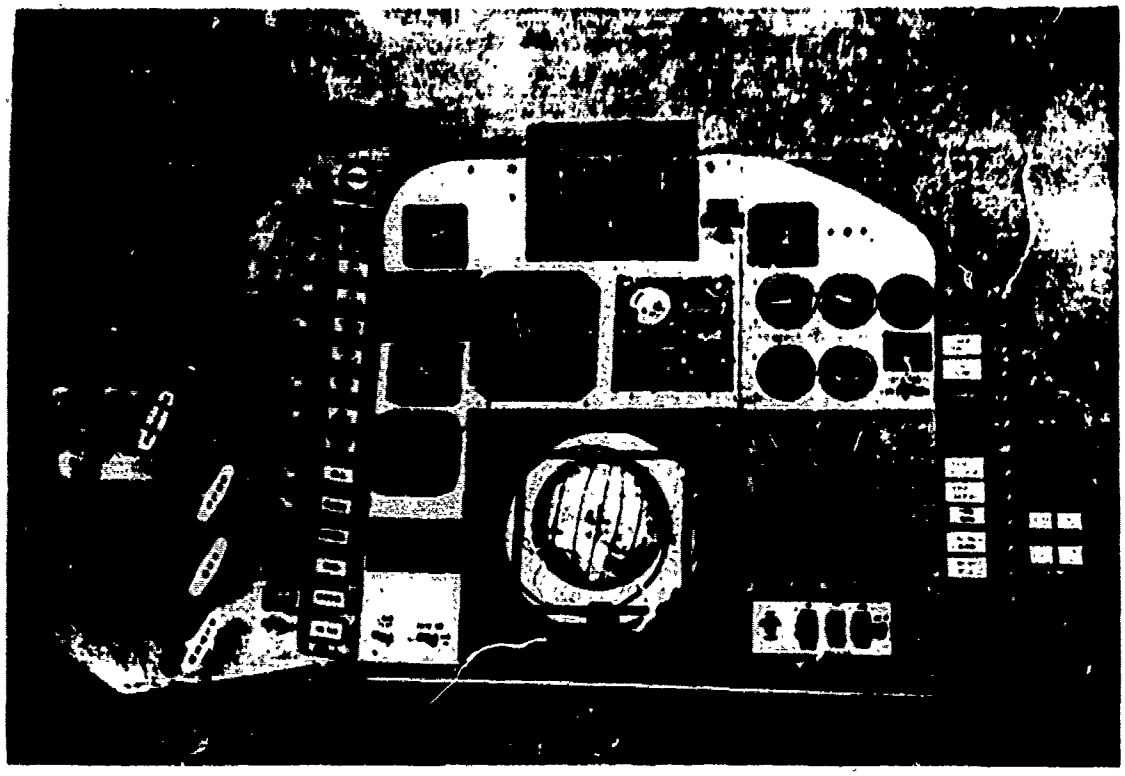


Fig. 17 Mercury type instrument panel

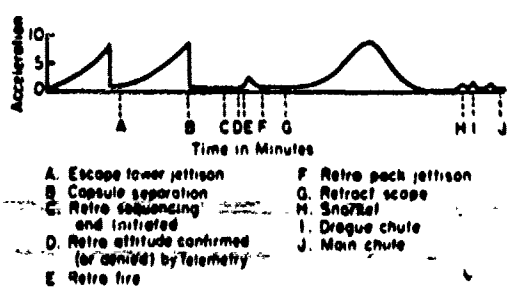


Fig. 18 Approximate G profile with event times and associated telepanel indicators

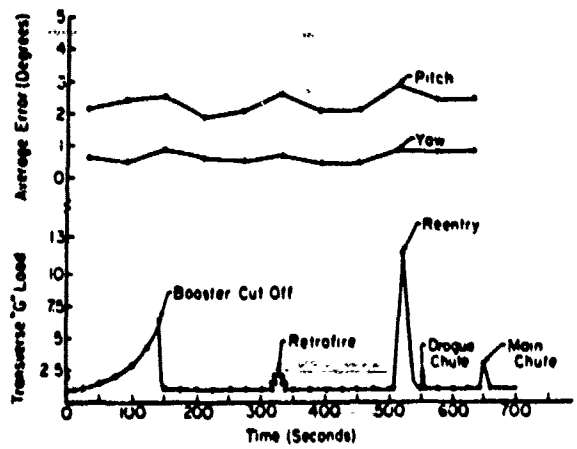


Fig. 19 Average pitch and yaw error correlated with Redstone G profile, for astronauts A, B, and C

also became skilled in the operation of their environmental control systems and capsule communication procedures during acceleration exposure. Simultaneously, extensive physiological monitoring and performance provided continuous information on astronaut endurance and piloting skill. Figure 19 shows an example of a typical Redstone acceleration profile during which time the ability of the astronauts to track in pitch and yaw dimensions was monitored. The figure shows that there was very little effect of the acceleration on their tracking skill during these particular runs.

"Complete mission simulations" were presented during which early morning suiting, psychiatric testing, waiting in the gondola, launch, orbit, reentry, escape, post-flight testing, and debriefing were provided on a real-time basis. This type of simulation presented physiological and psychological conditioning and man-machine evaluations along real-time scale profiles. An additional advantage of this type of training was that the astronauts were able to experience the many subtle and elusive interactions which occur between the physiological, psychological and engineering stress variables. Evaluations of the AMAL centrifuge as a training device have been very favorable (Slayton, 24; Glenn, 25, 26; Shepard, 27; Grissom, 28; Voas, et al, 29; Voas, 30; Carpenter, 31.)

VISUAL PERFORMANCE

At a recent study conducted at AMAL in cooperation with the Cornell Aeronautical Laboratories, it was demonstrated that the minimally acceptable (threshold) contrast was greater for positive acceleration than for transverse acceleration. In this particular experiment, visual brightness discrimination was studied at four levels of background luminance, at four levels of positive acceleration, and at five levels of transverse acceleration. In this study, the subject in the centrifuge viewed a circular test patch against a diffuse background, and operated a response button to indicate the appearance or disappearance of the test patch. Each centrifuge test run provided approximately 15 responses, each of which was at peak G for 90 seconds. Using six healthy subjects with 20/20 vision, brightness discrimination thresholds were determined at positive acceleration levels of 1, 2, 3, and 5 G_z, and at transverse acceleration levels of 1, 2, 3, 5, and 7 G_x. Determinations were made at each G level with background luminance of .03, .29, 2.9, and 31.2 foot-lamberts. Figure 20 shows the observed relationship between brightness discrimination threshold and background luminance for each of the four levels of positive

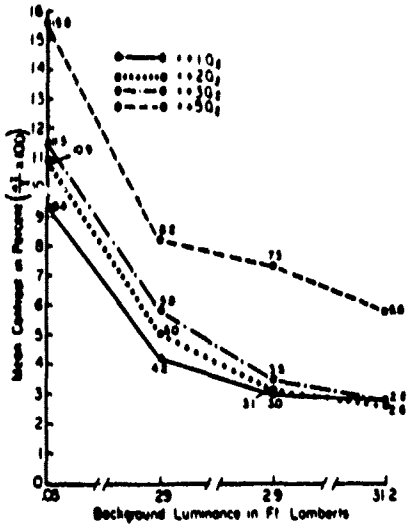


Fig. 20 Results of experiment showing relationship between brightness discrimination threshold and background luminance for four levels of positive acceleration

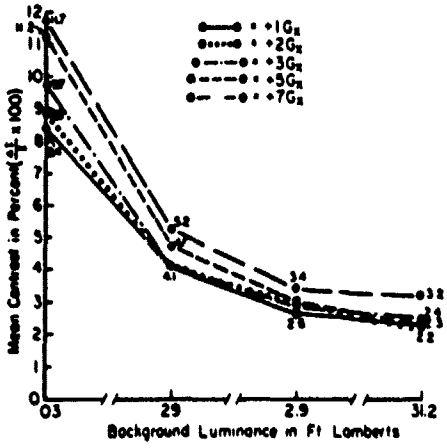


Fig. 21 Relationships between brightness discrimination threshold and background luminance for each of five levels of transverse acceleration

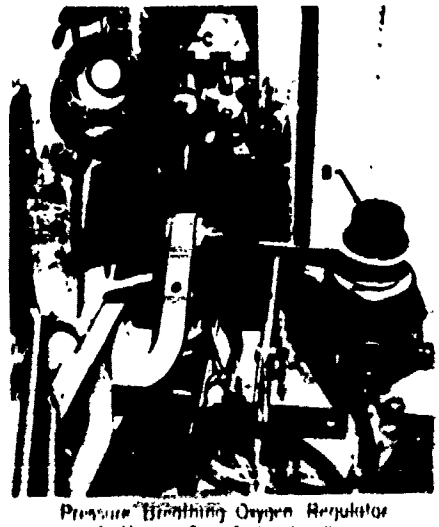


Fig. 22 Pressure breathing oxygen regulator installed in centrifuge

acceleration. Figure 21 shows the obtained relationships between brightness discrimination threshold and background luminance for each of the five levels of transverse acceleration. These figures show that for each of the four positive acceleration conditions the visual contrast requirements increased as the background luminance decreased. Also, for any given background luminance level, the higher acceleration levels required more brightness contrast. Similar results were shown for the transverse G exposures. A comparison of these two figures shows that positive acceleration stress consistently imposed higher contrast requirements than did transverse acceleration.

Armstrong (32) and Watson and Cherniak (33) have suggested that providing a pilot with positive pressure breathing of 100% oxygen during acceleration stress, especially sustained transverse acceleration, increases endurance time. Watson and Cherniak (33) reported a 67% increase in endurance time for subjects exposed to +10 G_x when positive pressure breathing of 100% oxygen was provided. Using the same visual brightness discrimination apparatus as described in the above section, an experiment was conducted at AMAL to determine whether positive pressure breathing of 100% oxygen would facilitate brightness discrimination during steady-state accelerations. The subjects operated a pressure breathing oxygen regulator manually (see Figure 22) so as to provide 0.7 inches of mercury per G conditions of both transverse acceleration and positive acceleration exposures on the centrifuge. The subjects performed under three breathing conditions: breathing normal air, 100% oxygen, and 100% oxygen under positive pressure. Given a background luminance of .03 foot-lamberts, the subjects were required to repetitively operate a switch (see Figure 22) to maintain the target at the minimally discriminable brightness contrast level. The results obtained under positive pressure (G_z) acceleration exposures of 90 seconds duration, each under acceleration loads of 1, 3, 4, and 5 G are shown in Figure 23. These data suggest that at the 3, 4, and 5 G_z levels, the positive pressure plus 100% oxygen required less visual contrast than was required under the other experimental conditions. Similar results are shown for G_x accelerations, as shown in Figure 24. The contrast required for discrimination appeared to be the same for both the 100% oxygen and 100% oxygen plus positive pressure breathing. It is interesting to note that as acceleration increased, the percentage of subjects reporting beneficial effects from the positive pressure breathing of 100% oxygen increased, as compared with the other conditions (see Figure 25). (The details of this study are reported by Chambers, et al, 34).

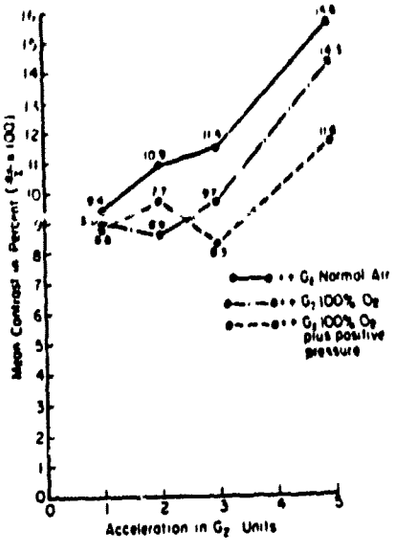


Fig. 23 Mean visual brightness contrast in percent versus acceleration in G_z units for normal air, 100% O_2, and 100% oxygen plus positive pressure

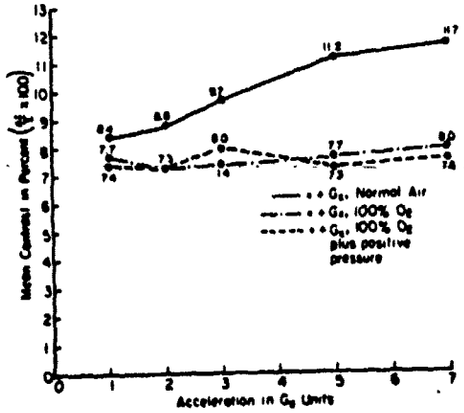


Fig. 24 Comparison of effects of breathing normal air, 100% oxygen, and 100% oxygen plus positive pressure on brightness contrast requirements

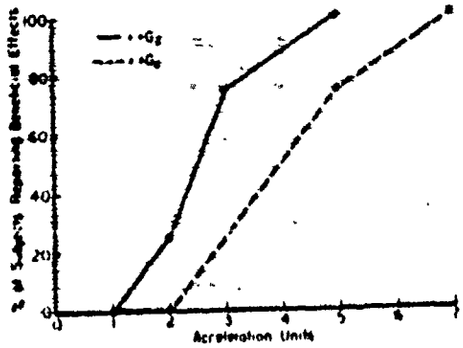


Fig. 25 Percent of subjects reporting beneficial effects of positive pressure breathing of 100% oxygen as compared with normal breathing of 100% oxygen and normal air

A similar study was conducted on test pilots at much higher acceleration loads, using peak acceleration centrifuge runs of 8, 10, and 12 G's. The pilots were volunteers from the USAF Aerospace Pilot's School at Edwards Air Force Base. The pilots performed a Mercury type reentry task, with the centrifuge at a steady-state acceleration level for 2 minutes under each breathing condition. The centrifuge installation for this experiment is shown in Figure 26. At +12 G_x, there were definite suggestions that performance under conditions of positive pressure breathing of 100% oxygen was superior to normal atmospheric breathing of 100% oxygen. Figure 27 summarizes these data. At 6, 8, and 10 G_x, however, no differences were observed. Subjectively, the pilots reported that positive pressure breathing of 100% oxygen was superior to the condition of normal breathing of 100% oxygen in terms of breathing ease and general comfort (Chambers, et al, 34).

RESPONSE TIME

Although it is generally agreed that acceleration influences discrimination reaction time behavior, it has not been possible to identify all of the underlying mechanisms which mediate these effects. During acceleration, the changes observed in reaction time may be associated with pilot impairment in a variety of physical loci. Acceleration might reduce the capacity of the peripheral system to receive the stimulus, or of the central nervous system to process already received stimuli and to indicate discriminatory choice, as well as reduce the ability of the neuromuscular system to coordinate the motor components which translate the response into the manipulation of the appropriate control device. In addition, some studies have indicated that discrimination reaction time under G is indirectly affected by the protective equipment and related components present in the situation in which the tests are conducted.

For example, in Figure 28 it is difficult to determine which way, if any, suit conditions during dynamic simulation influenced reaction time to telepanel light indications. Similarly, it is difficult to determine whether acceleration conditions influenced these reaction times. Overall mean response times to telepanel indications were not significantly affected. However, detailed analysis of the scores around any given point indicate that the response variability demonstrated by the astronauts was significantly increased by the acceleration condition, and by the suit hard condition. Similarly, in a later study involving these same astronauts, as is shown in Figure 29, the average response time for the suit hard condition was slightly greater than for the suit soft condition.



Fig. 26 Pilot, mask, restraint system, three-axis side-arm controller, and instrument display panel used in positive pressure breathing experiment

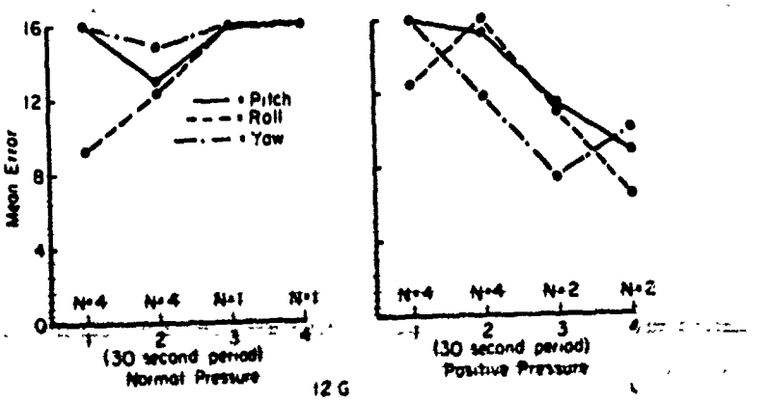


Fig. 27 Mean degrees error for pilots in pitch, roll, and yaw for four successive 30-second intervals during exposure to 12 G_x under conditions of positive pressure breathing of 100% oxygen, and normal breathing of 100% oxygen

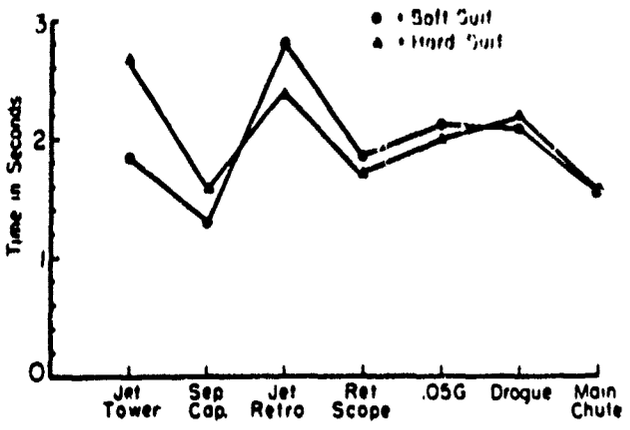


Fig. 28 Mean reaction times to seven telelights in ~~Mercury-III centrifuge~~ simulated flight profile

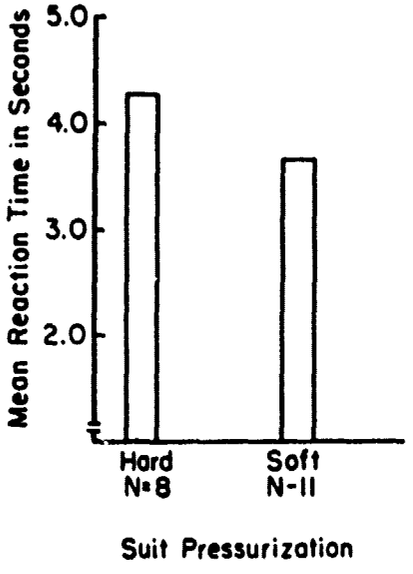


Fig. 29 Telepanel response time as a function of suit conditions

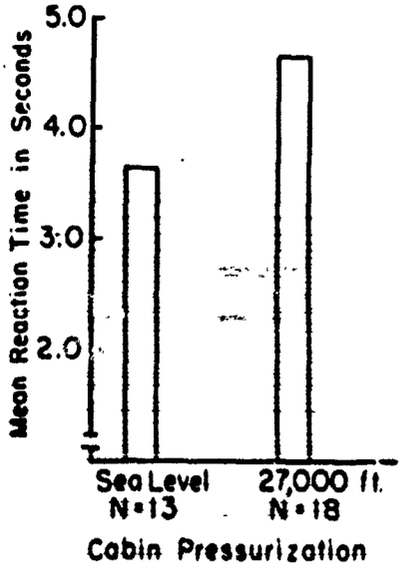


Fig. 30 Telepanel response time as a function of cabin pressure

The average latency was 4.28 under suit-hard conditions, and 3.68 seconds during the suit-soft conditions. These differences were not statistically different. It is interesting to note that the average response times were not significantly affected by the change in altitude simulated by gondola evacuation, as shown in Figure 30. Here, the average override latency was 3.64 seconds at sea level pressure (14.7 psi) and 4.68 seconds when the gondola pressure was reduced to 5 psi.

The above studies were all conducted in relatively complex Mercury simulation projects in which the pilots were performing rather complex tasks during complete mission acceleration conditions. However, in a series of basic research studies, an attempt was made to measure the effects of steady-state transverse acceleration on discrimination time without the involvement of pressure suits and gondola evacuation. A discrimination reaction time test apparatus was developed that consisted of four small stimulus lights, a small response handle containing four small response buttons, and a programmer device which could present a large variety of random sequences to subjects on the centrifuge (Chambers, Morway, et al, 10). As each of the lights came on, the subject was required to press the associated finger button with his right hand as fast as he could. Both the automatic program which activated the stimulus lights and the subject's responses were fed to an analog computer where initial data reduction was accomplished. Following pre-acceleration training to establish a stable baseline performance level, each subject received three blocks of 25 trials each while exposed to +6 G_x for five minutes. Each subject received three such acceleration trials. Since speed and accuracy were both involved in this type of response behavior, times and errors were normalized and added. The results are shown in Figure 31 in which a highly significant effect of the acceleration on performance was demonstrated. Further, the effect persisted to the post-test period.

Using an auditory task rather than a visual stimulus in order to avoid the problem of visual interference which accompanies acceleration, it has been possible to obtain data on auditory reaction time at grayout levels. One such task (Cope and Jensen, 35) required the subject to add pairs of numbers which he heard via an auditory magnetic tape system and then to describe the sum by pressing the small odd and even response buttons which were mounted upon his left and right hand grips, respectively. Research with this apparatus during positive acceleration exposures to grayout levels indicated that the time required to make these responses increased during exposure to positive acceleration.

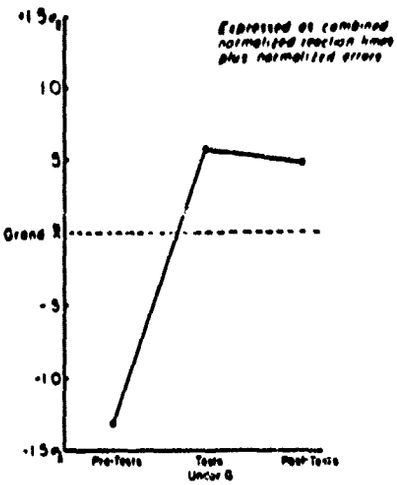


Fig. 31 Complex discrimination reaction time performance measures recorded under sustained -8 G_x

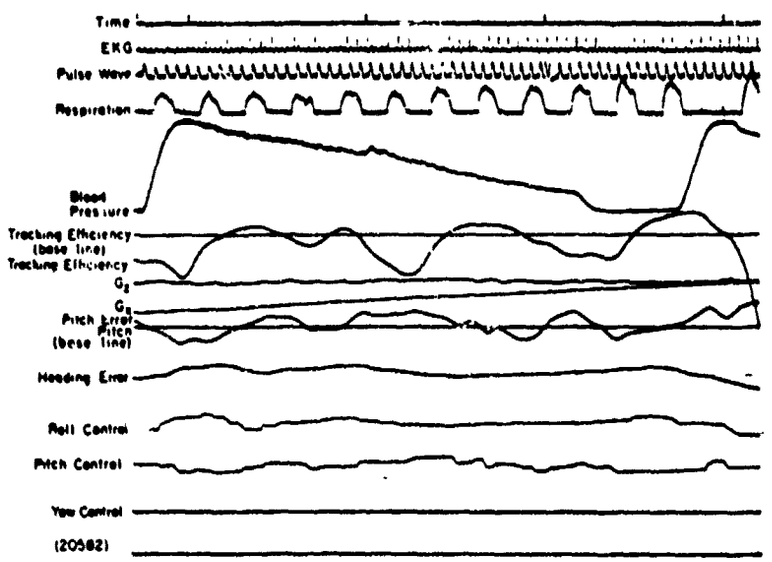


Fig. 32 Example of physiological and performance measures recorded under sustained -8 G_x

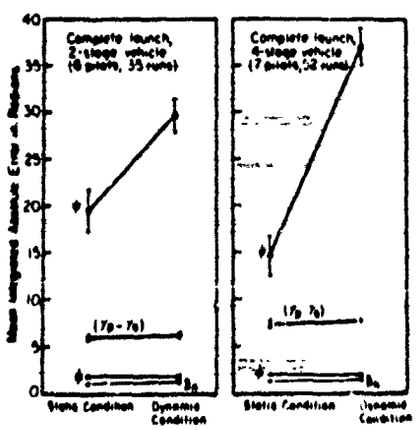


Fig. 33 Static versus dynamic conditions for 2-stage and 4-stage vehicles

COMPLEX SKILLS

When pilots are trained to perform complex psychomotor tasks, and then subjected to stressful steady-state acceleration runs on the human centrifuge, the effect of the acceleration on skill proficiency may be readily demonstrated. Skill decrement usually occurs prior to physiological decrement. For example, Figure 32 shows EKG, pulse, respiration, blood pressure, tracking efficiency, pitch error, heading error, roll control, pitch control, and yaw control at -8 G_x centrifuge run. In this particular example, tracking efficiency was calculated in percentage units based on accumulated tracking error divided by the accumulated excursion of the target display which the pilot was monitoring. Pitch and roll control inputs were made with a small pencil controller, and proficiency could range on a percentage scale from 100% to -100%, as derived from the division of the actual control output by the required output. This figure clearly shows that the tracking efficiency took a very sudden and marked drop from nearly 90% to approximately -95% near the end of the run. Very little physiological change is shown except for a slight change in respiration. This record is one of the many instances which have emphasized the predictive value of performance scores for medical monitoring purposes, and it illustrates the detrimental effects of high sustained acceleration on psychomotor skill performance.

Studies of pilot performance during staging acceleration profiles, such as may be characteristic of certain two-stage and four-stage launch vehicles, have suggested that at the higher acceleration levels, pilots find it extremely difficult to concentrate on all aspects of a complex task while they are exposed to high acceleration loads, whereas at the lower acceleration levels, they can perform very well. Figure 33 presents examples of this condition, in which pilots performed exactly the same tasks statically and dynamically for each of two types of booster combinations. The pilot's task was to perform the four aspects of the task continuously so as to fly the vehicle through the orbital injection "window". For both types of vehicles, it was found that the pilots made significantly more errors on the yaw quantity during dynamic conditions than during static conditions, but that they were able to maintain the other three task components very well under both dynamic and static conditions. In this particular study, the accelerations did not exceed +7 G_x for either type of vehicle.

It is interesting to compare the performance of a single astronaut who performs the same task on the centrifuge both

statically and dynamically. Figure 34 presents such a comparison. The most marked difference is noted between the static and dynamic condition during which times the astronaut was performing the re-entry task. Some impairment of performance is shown, and this is attributed to the acceleration.

During reentry simulations of the Atlas vehicle on the human centrifuge, inadvertent control inputs are not uncommon. These inadvertent inputs often mirror the acceleration profile under which a control task is being performed. Figure 35 shows an actual record of this, illustrating inadvertent control inputs in pitch. Figure 36 illustrates inadvertent control inputs in the roll and yaw axes, using a Mercury-type controller.

In addition to inadvertent inputs which accompany acceleration, other more general effects of dynamic conditions may be observed. Acceleration appears to generally reduce the sensitivity and timing of all controller movements. Figures 37 and 38 are sample portions of the recorded static and dynamic performance of the same astronaut taken within twenty minutes of each other. These records serve to illustrate the general effects of acceleration upon the frequency and amplitude of control movements.

During exposure to high transverse accelerations, the effects of the duration variable on pilot performance have been demonstrated recently in an experiment conducted on nine Aerospace Research Pilot School students. At 15 G_x, for example, most of the pilots were able to perform a tracking task for approximately two seconds before major performance decrement occurred. Typical results, as obtained from one pilot, are shown in Figure 39, in which major performance decrement is shown during the last half of the 15 G_x run. In contrast with this, the pilot's performance at 10 G_x for 30 seconds remained consistently good throughout. Similarly, at 12 G_x, during exposures of 15 seconds, the pilots typically showed major decrement beginning at the middle of the run, e.g., after approximately 7 1/2 seconds at 12 G_x. Performance during the first half was good on the average. This effect, generally attributed to fatigue, is illustrated in Figure 40 for two pilots at 12 G_x.

HIGHER MENTAL SKILLS

In cooperation with Rutgers University, we developed a continuous memory testing apparatus which could be used on the AMAL centrifuge. It required the continuous and repetitive

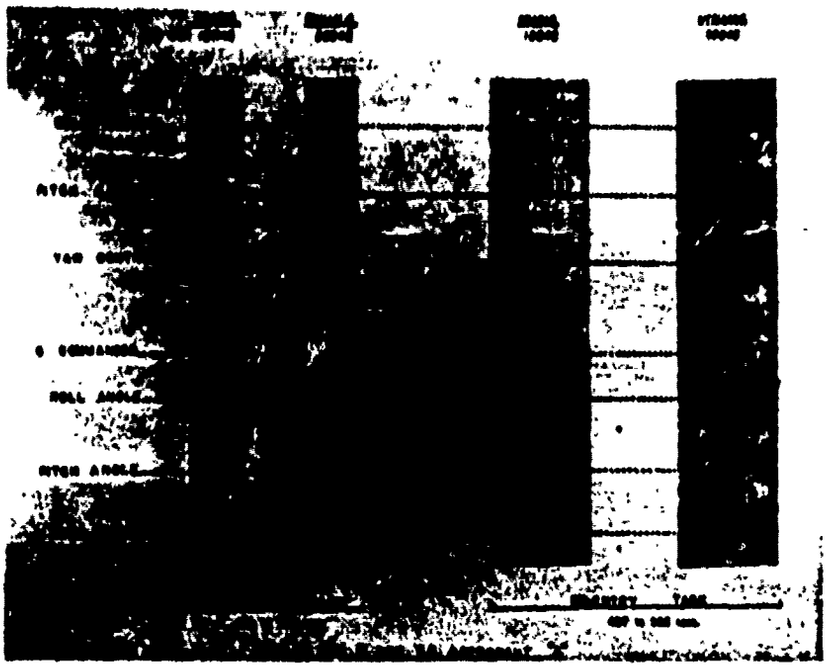


Fig. 34 Sample recordings of Astronaut A's performance on retro-fire versus reentry profiles, static and dynamic

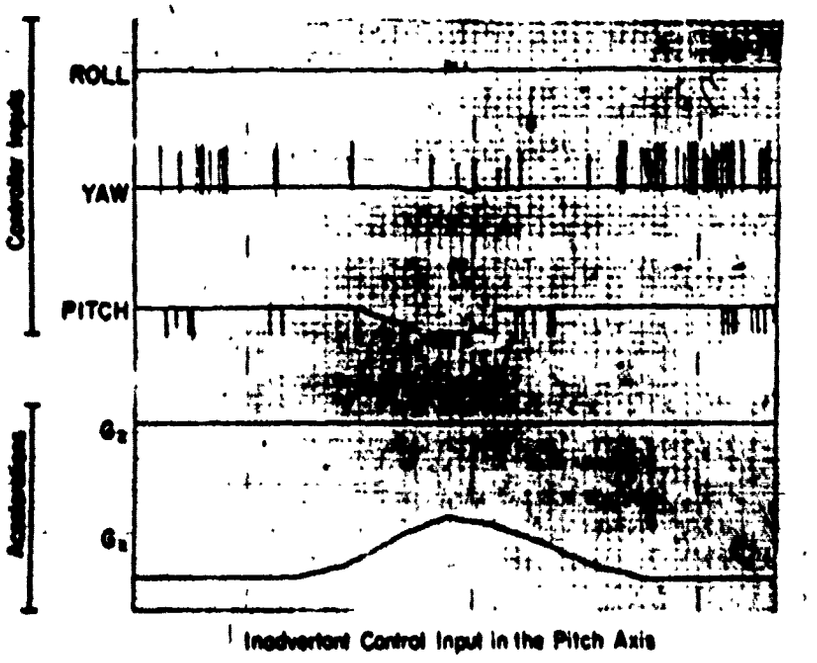


Fig. 35 Recording of inadvertant control input in the pitch axis

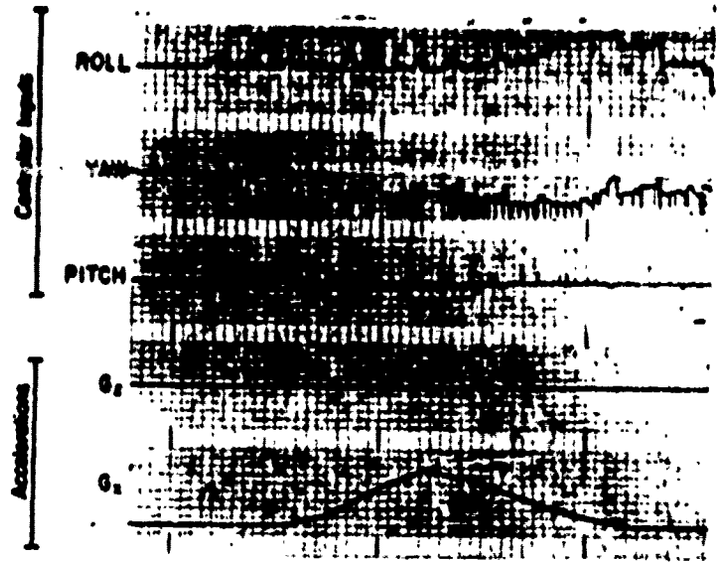


Fig. 36 Recording of simultaneous inadvertant inputs in the roll and yaw axes

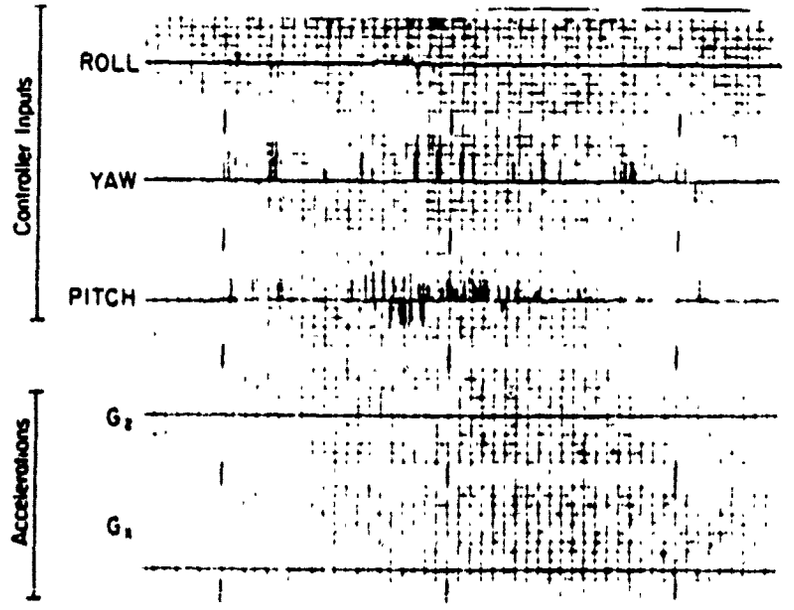


Fig. 37 Recording of sample performance under static conditions

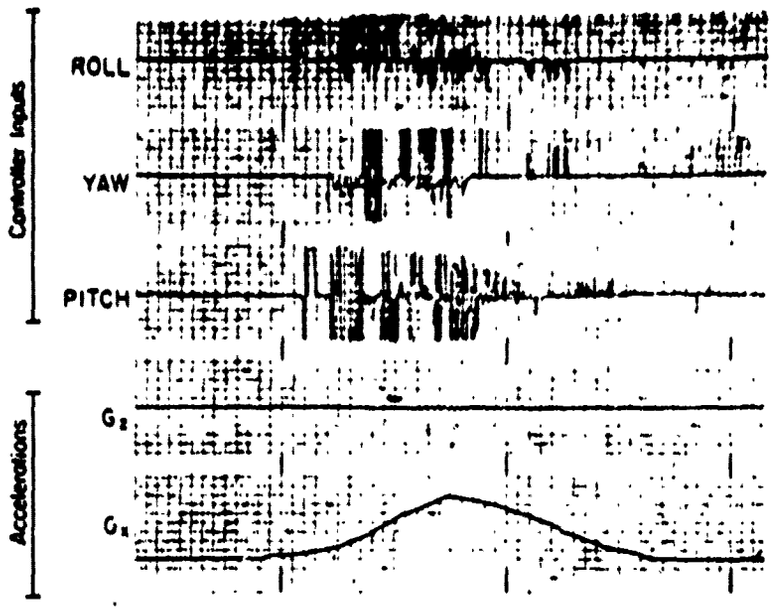


Fig. 38 Recording of performance of the same pilot under dynamic conditions

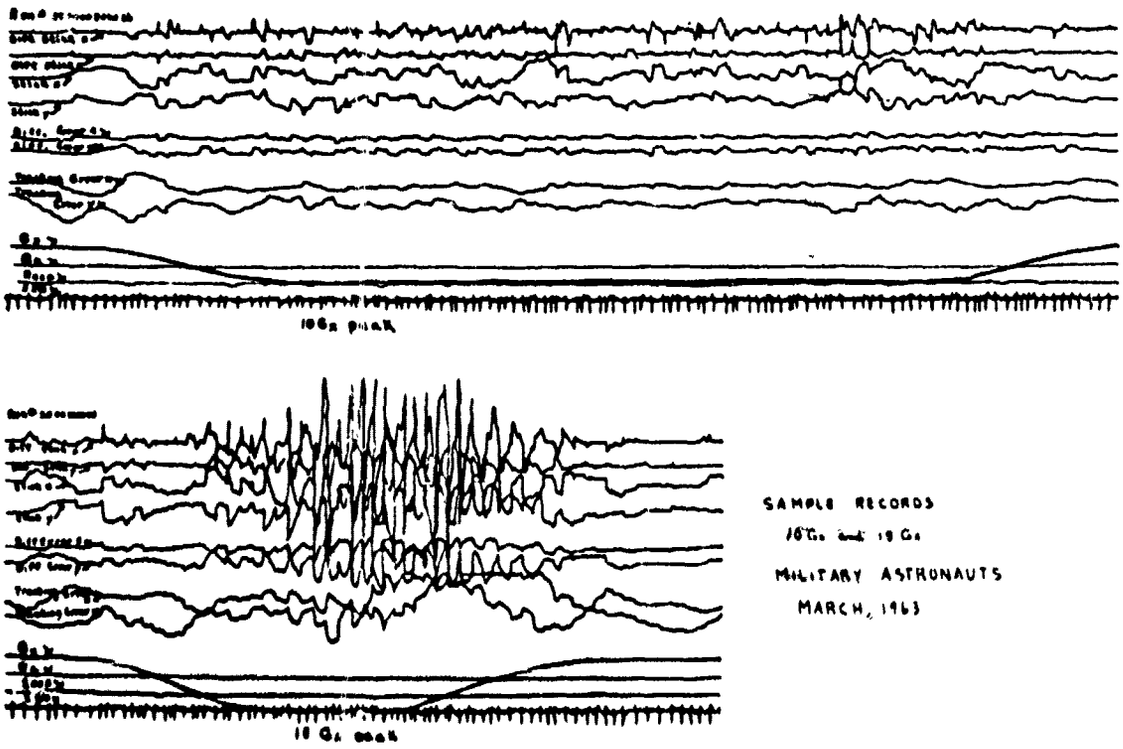


Fig. 39 Sample recordings of piloting performance during exposure to 10 and 15 G_x on the human centrifuge

memorization of a portion of a sequence of random symbols. As each symbol occurred, the subject was required to compare it with his memory of the symbol that had been presented to him two, three, or four presentations previously. New symbols appeared continuously so that the subject constantly had to forget earlier symbols as he added new ones. Approximately 50 symbols were presented for each of the runs. In the earlier study, each run stayed at +5 G_x for five minutes. The data, collected on 21 subjects, indicated that the subjects could continue to perform this task just as well during exposure to +5 G as they could statically. However, subjectively, the subjects reported that their performance deteriorated under G and that they generally regarded +5 G_x for five minutes as a stressful experience. It was interesting to note, however, that subjects tested in the centrifuge gondola did more poorly than subjects tested in a regular testing room. The implication seems to be that some apprehension or anxiety may have been acting to interfere with maximum performance. The results of this experiment are shown in Figure 41.

In a more recent study conducted on the AMAL centrifuge, we developed a task which required the subject to monitor two small display tubes which were located directly in front of his normal line of vision. The left-side tube presented numbers, and the right-side tube presented plus and minus symbols. The task was to continuously make matches for these two presentations simultaneously as the runs proceeded and to select one of two buttons to indicate whether both the number and symbol which were then appearing were the same as or different from those which had occurred on a specified number of trials previously. Acceleration loads of 1, 3, 5, 7, and 9 G's were studied. Each test was 2 minutes and 18 seconds long. The results of the experiment suggested that proficiency in immediate memory was maintained at least through 5 transverse G. However, at +7 and +9 G_x, some impairment of immediate memory was observed. The results of this experiment are shown in Figure 42.

EFFECTS OF PROLONGED LOW-G ACCELERATION

During prolonged exposure to acceleration, the continuous concentration necessary for performance is difficult, fatiguing and boring. For example, during an extended 2 G centrifuge run which lasted 24 hours, the subject started out with a somewhat detailed set of procedures to follow in making medical observations upon himself, recording his subjective comments, and writing and typing (Hardy, et al, 23). However, the subject found that in spite of his

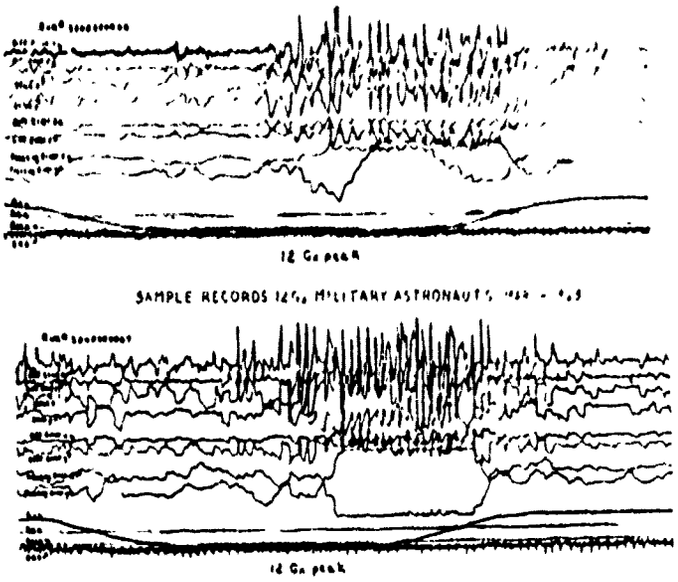


Fig. 40 Sample recordings of piloting performance during exposure to 12 G_x on the human centrifuge

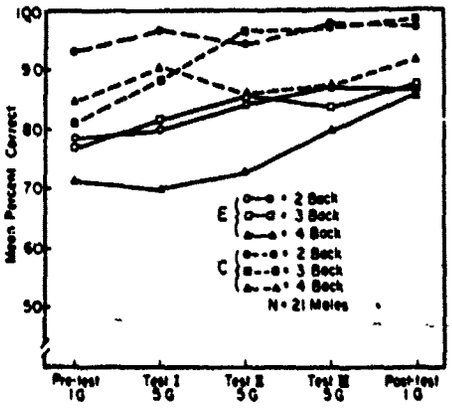


Fig. 41 Mean percent correct for static, 5 G, and post-test runs, control versus experimental groups, on 2-back, 3-back, and 4-back immediate memory task

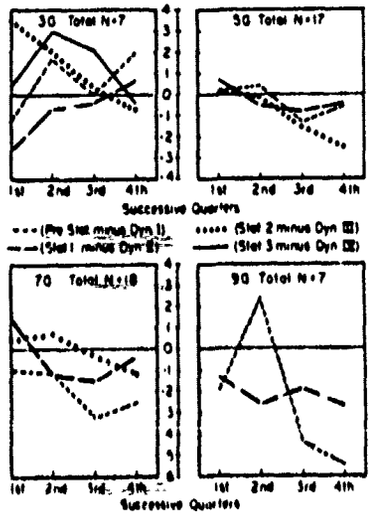


Fig. 42 Successive quarters scores for 3 G, 5 G, 7 G, and 9 G_x, showing performance on immediate memory task

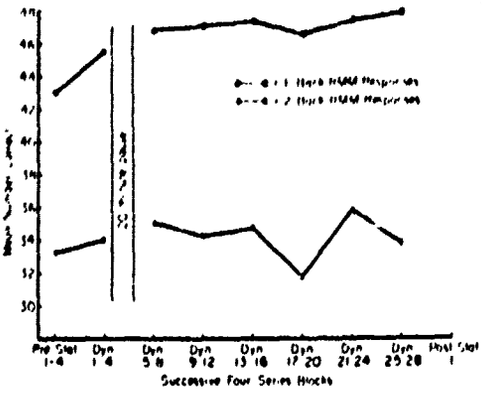
initial high resolve, he took naps and listened to the radio and suffered primarily from boredom and fatigue. Areas of contact with the chair in which he was seated were the sources of the greatest localized discomfort. At 16 hours elapsed time, the subject reported the onset of aesthenia of the ring and little finger and outer edge of the palm of the left hand. The subject found it impossible to maintain his originally prescribed maintenance and observation schedules.

Chambers and Ross secured a subject in a Mercury-type contour couch and required him to perform the two-symbol running matching memory task (previously described) every 10 minutes for four and one-half hours. The subject was able to perform this task throughout the entire period with only minor performance impairment. The results of this experiment are shown in Figure 43.

INTERACTION EFFECTS

It has been shown that the human centrifuge is a useful tool for measuring the interaction effects of several variables simultaneously. In addition to both the direct effects of acceleration upon human performance and the less obvious interactions between performance and acceleration already mentioned, there is a growing body of information pertaining to the somewhat secondary role that other flight conditions play in determining a pilot's performance during acceleration stress. For example, in the early simulations of proposed space vehicles, several types of right hand side-arm controllers were tested on the AMAL centrifuge (Chambers 6). Figure 44 presents a diagram of four of these controllers: a three axis balanced controller with all three axes intersecting; a three axis controller (unbalanced) having none of these three intersecting; and a finger-tip controller having two intersecting axes (and yaw for toe pedal operation); and a two axis controller with axes which do not intersect (coupled with toe pedals for yaw control).

In Figure 45 the effects of two specific acceleration fields upon pilot performance during the pitch and roll maneuvers involved in a tracking task are shown for each of the four types of controllers. When the pilots performed in one acceleration field, their error performance on all four controllers was essentially the same. However, when these same pilots flew the same problem under a different acceleration performance on Type II controller deteriorated while performance on the other two remained essentially the same. A similar change in G field resulted in an increment in error for Types II and III controllers and reduction in error



Successive Four Series Blocks

Fig. 43 Mean number correct as a function of successive four-series blocks for 1-back and 2-back immediate memory task responses during a 4-1/2 hour centrifuge run at +2 G_x

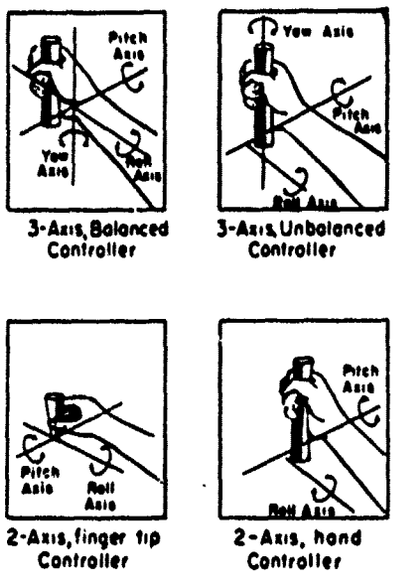


Fig. 44 Controllers of various types – 3-axis balanced, 3-axis unbalanced, 2-axis finger tip, and 2-axis hand control

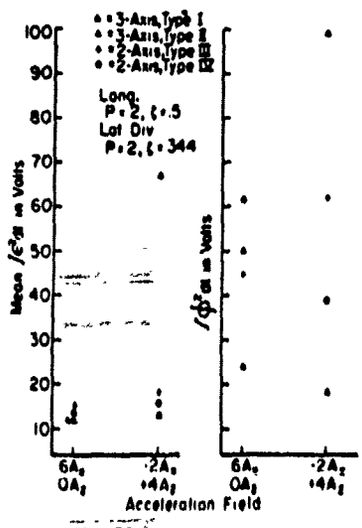


Fig. 45 Performance error in various acceleration fields as a function of the type controller used

for Type IV, resulting in a shift in rank order of the controllers. The differential effects upon performance induced by different types of acceleration on controllers are shown in Figure 46. Here the mean tracking proficiency score for the test pilots who performed the same tracking tasks using each of the four different types of side-arm controllers within given acceleration fields and under varying amounts of cross coupling and damping are shown. This figure shows not only the effect of using different specific G fields on particular tracking tasks, but also illustrates the effects of damping and cross coupling when the effects of acceleration are held constant.

A similar type of result may be shown which resulted from some of the Mercury astronaut training programs on the AMAL centrifuge. In some of these simulations it was observed that the tendency to use less discrete, more frequent control inputs under dynamic conditions was associated with an overall increase in fuel utilization. A most important aspect of this relationship depends upon the fact that differential rates of fuel usage were observed even when no significant differences in adequacy of control as measured by integrated attitude error were present. As previously indicated, pilot ability to damp the reentry oscillations in pitch and yaw was reduced under dynamic simulation. In contrast, control capability in the roll axis was not significantly affected by dynamic reentry accelerations. Therefore, roll control during reentry can be used to illustrate this dynamic effect under conditions of equivalent error. Figure 47 illustrates not only the correlation between incurred roll rate error and compensatory fuel usage, but also that fuel utilization was approximately 33% greater under dynamic than under static conditions, though integrated error was of the same approximate magnitude under both conditions. Data of this nature emphasize the advisability of obtaining both dynamic and static performance before placing estimated values upon such design parameters as required fuel reserves.

Another example of interaction effects regarding pilot performance is shown in Figure 48. As may be seen in this figure, the hard suit (5 psi differential pressure) conditions resulted in a reduction in relative piloting performance as measured by the percentage of reentry simulations in which capsule oscillations were successfully damped during static simulation of the reentry control task, but appeared to assist performance under dynamic conditions. The performance values presented in this figure are not absolute but they represent relative performance using the conditions of static soft-suit, under which control must often be retained throughout

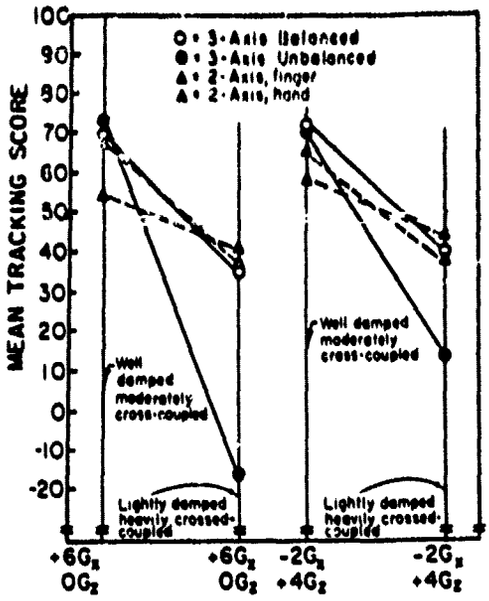


Fig. 46 Mean tracking score for 3-axis balanced, 3-axis unbalanced, and 2-axis finger and 2-axis hand controllers

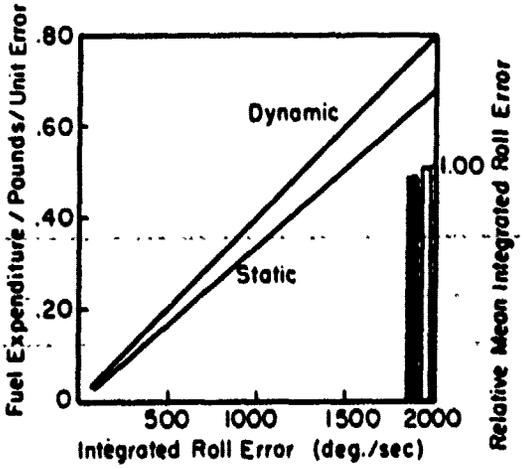


Fig. 47 Fuel utilization as a function of error magnitude and acceleration

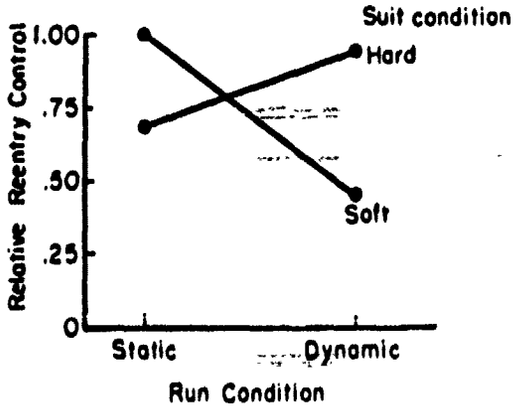


Fig. 48 Relative reentry control as a function of suit pressurization and acceleration

the reentry profile, as a base-line referent. The additional forearm support provided by the pressurized suit appeared to reduce the frequency and/or magnitude of the previously described inadvertent inputs which accompanied dynamic simulation. As the tendency to insert such inputs was reduced through practice, the stabilization provided by the inflated suit appeared to become less and less of an advantage and the interaction between suit and run conditions was markedly less during the latter stages of training. Verbal reports obtained toward the end of the training program indicated that the pilots considered suit-hard conditions more uncomfortable and perhaps even less effective.

PERSONALITY, MOTIVATION, AND EMOTIONAL BEHAVIOR

The effects of acceleration on personality, emotional behavior, and motivation cannot be measured as easily as the more quantitative performance aspects which have been presented in earlier sections of this report. Most of the observations to date are based on the verbal comments of subjects during acceleration exposures, debriefing sessions, films, television, the results of interviews and questionnaires, and physiological measures taken during the acceleration stress. It is difficult to delineate the personality, emotional, and motivational aspects because of: (1) large individual differences, (2) lack of quantitative measures which may be administered during the run, and (3) the necessary priority physical safety exerts, (making difficult the administration of the currently available psychological instruments). It is a common observation in acceleration laboratories, however, that both the anticipation and occurrence of acceleration forces contribute to anxiety and other types of motivational behavior.

At AMAL, the Mercury Astronauts and other volunteer subjects have been exposed to as much as 8 hours confinement in the centrifuge, under conditions in which various acceleration stresses were imposed. In these studies, the astronauts were confined within their pressure suits, strapped in their contour couches, placed within the closed gondola of the centrifuge, and presented with realistic pre-flight tasks, monitoring requirements, and communications. After one or two hours pre-launch wait, they went through accelerations expected on the Atlas rocket, and spent approximately 4 1/2 hours in simulated orbit. This was followed by reentry accelerations and egress. In some of these studies, the launch and reentry acceleration profiles were simulated successively within a single all-day session. Whereas extensive personality, biochemical, performance, and physiological tests have been

conducted, the results in general have not shown any significant long-term personality or emotional effects. In most instances, it was not possible to make quantitative measures of personality, emotional behavior, or motivation immediately before, during, or after these centrifuge exposures because of the extensive number of other engineering and physiological tests which were required and also, there are no good quantitative measures of personality. However, pilots and astronauts have frequently reported that variables other than the acceleration itself were more stressful. Pressure points, muscle cramps and pains due to the pilot's restraint system, uncomfortable temperature, failure of the urinary bag, discomfort due to the biomedical sensors or special flight gear, and long-term delay in planned testing due to equipment malfunction, have created extremely aggravating and sometimes emotionally disturbing conditions.

All available data to date seem to indicate that nearly all of the emotional, personality, and motivational aspects are highly susceptible to training and experience. It appears that the well-trained subject is capable of compensating for and overcoming many psychological problems which he may encounter before, during or following high-G acceleration exposures, whether they be on a centrifuge, a rocket sled or in a spacecraft.

SUMMARY

This paper has reviewed some of the major effects of long-term acceleration on human performance and physiology, and has presented some of the results of recent centrifuge simulation studies of manned space flights and proposed spacecraft. The paper has been primarily concerned with the results and studies conducted on the AMAL Human Centrifuge in support of NASA and USAF space projects, and acknowledgment is made of the contributions of the many pilots, astronauts, scientists, and engineers who participated in these projects. Physiological tolerance to $+G_x$, $+G_z$, $-G_x$, and $-G_z$ acceleration vectors are discussed, and it is pointed out that in addition to physiological tolerance limits, there are also performance tolerance limits which define the reliable functioning of any particular overt behavior system during acceleration. Major concepts in protection are discussed, recent work on contour couches studied on the AMAL centrifuge is reviewed, and a newly developed universal contour couch is described. Some of the basic principles of centrifuge simulation of several types of space vehicles are outlined, and the use of the AMAL centrifuge in astronaut training is reviewed. The remainder of the paper is concerned with the effects of various acceleration profiles on visual performance,

response time, complex psychomotor skill performance, and higher mental functioning. Following a discussion on the effects of prolonged low-G acceleration exposure, some of the effects of combinations of variables such as acceleration, pressure suit conditions, and type of side arm controllers are reviewed. Some of the problems in measuring the effects of acceleration on personality, motivation, and emotional behavior are outlined, and the susceptibility of these factors to training and experience are emphasized. It is concluded that there are major deficits in our knowledge concerning the effects of complex acceleration profiles which are combined with other stress conditions encountered during flight.

REFERENCES

1. Gauer, O.H. & Zuidema, G.D. Gravitational Stress in Aerospace Medicine, Boston, Little, Brown & Company, 1961.

2. Price, J.F. Physiological & Psychological Effects of Space Flight. Vol. 1, Acceleration, Deceleration, and Impact. Space Technological Laboratories, Inc., Redondo Beach, California, October, 1962.

3. Wood, E.H. & Lindberg, E.F., Acceleration, In Physiology of Man in Space, pp 61-111. New York, Academic Press, 1963.

4. Brown, J.L., Acceleration and Motor Performance Human Factors, 2, pp 175-185, 1960.

5. Brown, J.L., The Physiology of Acceleration - Performance, pp 90-114, In, Gauer, O.H. & Zuidema, G.D. (Editors): Gravitational Stress in Aerospace Medicine. Boston: Little, Brown & Company, 1961.

6. Chambers, R.M., Control Performance Under Acceleration with Side-Arm Attitude Controllers, Aviation Medical Acceleration Laboratory, Johnsville, Pa., Report No. NADC-MA-6110, 27 November 1961.

7. Chambers, R.M., Human Performance in Acceleration Environments, In, Burns, N.M., Chambers, R.M., and Hendler, E. (Eds.) Unusual Environments and Human Behavior, New York: The Free Press, McMillan, 1963a.

8. Chambers, R.M., Psychological Aspects of Space Flight. In J.H.U. Brown (Ed.) The Physiology of Space Flight, New York: Academic Press, 1963b.

9. Chambers, R.M., and Doerfel, H.V., Closed-Loop Centrifuge Simulation of Space Vehicle Performance, Amer. Rocket Society, Semi-Annual Meeting, Preprint No. 807-59, June 8-11, 1959.

10. Chambers, R.M., Morway, D.A., Beckman, E.L., DeForest, R., Coburn, K.R., The Effects of Water Immersion on Performance Proficiency. Aviation Medical Acceleration Laboratory, Johnsville, Pa., NADC-MA-6133, 22 August 1961a.

11. Chambers, R.M. and Nelson, J.G., Pilot Performance Capabilities During Centrifuge Simulations of Boost and Reentry. American Rocket Society Journal, Vol. 31, No. 11, pp 1534-1541, Nov., 1961b.

12. Chambers, R.M. and Hitchcock, L., Effects of High G Conditions on Pilot Performance, pp 204-227. In, Proceedings of the IAS-NASA National Meeting on Manned Space Flight, Institute of the Aerospace Sciences, New York, May 1, 1962a.

13. Clark, C.C., Hardy, J.D. and Crosbie, R.J., A Proposed Physiological Acceleration Terminology with an Historical Review. Pages 7-65, In, Human Acceleration Studies, Publication 913, National Academy of Sciences - National Research Council, Washington, D.C., Panel on Acceleration Stress for the Armed Forces - NRC Committee on Bioastronautics, 1961b.

14. Dixon, F. and Patterson, J.L., Jr., Determination of Acceleration Forces Acting on Man in Flight and in the Human Centrifuge In, Gauer, O.H. & Zuidema, G.D. (Eds.) Gravitational stress in Aerospace Medicine, pp 243-256, Boston, Little, Brown and Co., 1961.

15. Gell, C.F. Table of equivalents for acceleration terminology. Recommended by the Committee on Acceleration of the AGARD Aerospace Medical Panel. Aerospace Med., 32, 1109-1111, 1961.

16. Eiband, A.M., Human Tolerance to Rapidly Applied Accelerations: A Summary of the Literature, National Aeronautics and Space Administration, NASA MEMO 5-19-59E, June, 1959.

17. Webb, M.G., Jr. End Points for Acceleration Tolerances on the Centrifuge, pp 59-64, In, Bio-Assay Techniques for Human Centrifuges and Physiological Effects of Acceleration, Ed. by P. Bergeret, New York, Pergamon Press, Agardograph No. 48, 1961.

18. Smedal, H.A., Vykukal, H.C., Gallant, R.P. and Stinnett, G.W., Crew Physical Support and Restraint in Advanced Manned Flight Systems. Amer. Rocket Society Journal, 31, 1544-1948, 1961.

19. Bondurant, S., Blanchard, W.G., Clarke, N.P. & Moore, F., Effect of Water Immersion on Human Tolerance to Forward and

Backward Acceleration, J. Aviat. Med., 29, 872-878, 1958.

20. Clark, C.C. and Gray, R.F., A Discussion of Restraint and Protection of the Human Experiencing the Smooth and Oscillating Acceleration of Proposed Space Vehicles. Aviation Medical Acceleration Laboratory, Johnsville, Pa., Report No. NADC-MA-5914, October, 1959.

21. Weaver, J. A., Rubenstein, M., Clark, C.C., and Gray, R.F., Encapsulation of Humans in Rigid Polyurethane Foam for use as a Restraint System in High Acceleration Environments, Aviation Medical Acceleration Laboratory, Johnsville, Pa., NADC-MA-6147, May, 1962.

22. Leverett, S.D.,Jr., Whitney, R.U. & Zuidema, G.D., Protective Devices Against Acceleration, pp 211-220, In. Gauer, O.H. and Zuidema, G. D. Gravitational Stress in Aerospace Medicine. Boston, Little, Brown & Company, 1961.

23. Hardy, J.D., Clark, C.C., Gray, R.F., Acceleration Problems in Space Flight, Aviation Medical Acceleration Laboratory, Johnsville, Pa. NADC-MA-5909, October, 1959.

24. Slayton, D.K., Pilot Training and Preflight Preparation, pp 53-60, In. Results of the First U.S. Manned Suborbital Space Flight, National Aeronautics and Space Administration, Washington, D.C., June 6, 1961.

25. Glenn, John H.,Jr. Readying the Mind and Body, pp 27-35 in Wainwright, L., (Ed.) The Astronauts: Pioneers in Space, N.Y., The Golden Press, 1961.

26. Glenn, John H. Jr. Pilot's Flight Report, pp 119-136, In., Results of the First United States Manned Orbital Space Flight, Manned Spacecraft Center, National Aeronautics and Space Administration, Houston, Texas, February 20, 1962.

27. Shepard, Alan B., Pilot's Flight Report, Including In-Flight Films, pp 69-75, In, Results of the First U.S. Manned Suborbital Space Flight, National Aeronautics and Space Administration, Washington, D.C., June 6, 1961.

28. Grissom, Virgil I., Pilot's Flight Report, In, Results of the Second U.S. Manned Suborbital Space Flight, Manned Spacecraft Center, National Aeronautics and Space Administration,

Houston, Texas, July 21, 1961.

29. Voas, Robert B., Van Bockel, John J., Zedekar, Raymond G., and Backer, Paul W., Results of In-Flight Pilot Performance, pp 61-67. In Proceedings of a Conference on Results of the First U.S. Manned Suborbital Space Flight. Washington, D.C.: U.S. Govt. Printing Office, 6 June 1961.

30. Voas, Robert B., Some Implications of Project Mercury, pp 41-44. In, The Training of Astronauts, National Academy of Sciences, National Research Council, Washington, D.C., 1961.

31. Carpenter, M.S., Pilot's Flight Report, pp 69-75, In. Results of the Second United States Manned Orbital Space Flight, NASA-SP-6, National Aeronautics and Space Administration, Manned Spacecraft Center, Houston, Texas, May 24, 1962.

32. Armstrong, R.C., The Effects of Positive Pressure Breathing on Transverse Acceleration Tolerance, Convair Aviation, Space and Radiation Med. Grp., San Diego, Rept. No. ZM-AM-001, 14 Jan 1959.

33. Watson, J.F., and Cherniack, N.S., Effect of Positive Pressure Breathing on the Respiratory Mechanics and Tolerance to Forward Acceleration, Aeronautical Systems Division, Aerospace Medical Laboratory, Wright-Patterson Air Force Base, Ohio, ASD TR 61-398, 1961.

34. Chambers, R.M., Kerr, Robert, Augerson, W.S., and Morway, D.A., Effects of Positive Pressure Breathing on Performance during Acceleration, Aviation Medical Acceleration Laboratory, Johnsville, Pa., NADC-MA-6205, 2 July 1962.

35. Cope, Freeman W., and Jensen, R.E., Preliminary Report on an Automated System for the Study of Mental Function in the Human Subjected to Acceleration Stress. Aviation Medical Acceleration Laboratory, Johnsville, Pa., NADC-MA-6113, 8 September 1961.

RESEARCH IN LIFE SUPPORT SYSTEMS FOR AIRBORNE AND SPACE VEHICLES

Roland A. Bosee
Captain, MSC, USN
Air Crew Equipment Laboratory
Naval Air Engineering Center
Philadelphia 12, Pa.

Optimal systems design in advanced aeronautics and astronautics demands the application of advanced technologies in many research areas. In some of these areas, the formulation of the problems requiring solution constitutes the primary problem. The identification and quantification of all the hazards threatening survival, safety, efficient performance, and comfort are not always possible until we have extended the bounds of existing knowledge. Empirical pioneering may be required. Two sets of mutually limiting variables must be determined, the physical, and the biological. They must then be synthesized through concurrent test, engineering development, and further research. Only in this way can inescapable environmental conditions be re-shaped for human tolerance, or at least toward that goal. This is the basic principle of our life support effort.

There is an important corollary to this principle. The problems must be formulated and the work must be planned, conducted and evaluated effectively. This requires workers who are able, willing and eager to digest new scientific information as an extension of existing knowledge. Should any serious lapse occur throughout the chain of necessary processes, an irretrievable and perhaps irreparable loss might result. And let me emphasize that, just as the best of advanced instrumentation often fails to yield what it should because of something like an unplugged cable, so may the most elaborate and costly efforts lead to confusion through the neglect of some simple detail. Ever since its establishment our Laboratory has experienced the need for preciseness in small things as well as in large ones. Nowadays that aviation medicine and space medicine must soar higher and faster than ever before to keep up with operational needs, scientific vigilance in our work has to be extended and intensified.

It is from the stated perspectives that several short episodes in the recent or current investigations at the Air Crew Equipment Laboratory (ACEL) have been selected for presentation at this Navy Science Symposium. They concern human engineering, altitude tolerance, atelectasis, and negative impact. In the short time available, I shall emphasize the broader conclusions derived from these examples, as they relate to difficulties or impending hazards in manned aircraft and space vehicles. The technical details have been omitted almost entirely, but I shall be happy to describe and discuss them later.

The first example deals with environmental protection. Excessive heat loads may be encountered for appreciable periods of time in high performance aircraft and spacecraft. A minimum of body clothing is therefore desired. However, the mission to be accomplished or some in-flight emergency may require the occasional use of a pressure suit. In order that the weight and volume of the vehicle be reduced as much as possible, we have been measuring the smallest geometric configuration in which a pressure suit can be donned and doffed. The tests are being run in our flexible cockpit simulator, shown in fig.1. The device consists essentially of a set of small, transparent planes, each one adjustable in both position and angle. While the subject goes through all the necessary motions with the suit, the dynamic spatial envelope is photographed against background grid lines in each of the three dimensions through fixed television cameras. With this system we can use a single focus for each camera and adjust the sizes of the three simultaneous images to a common scale electronically. The use of subjects of different percentile body dimensions, combined with the cranking of selected limitations to movement into the device permits a rational empirical analysis of the problem. The second figure is a view of the subject tying his boot laces. An acute end-point in these tests is the time taken to get into the suit, or the time to get out of it, without assistance. Another set of criteria has to do with the effort required. This is estimated in terms of measured changes of sweat rate, body temperature, and pulse rate. While the results of this small problem in human engineering may appear to border on the trivial, the study has demanded considerable ingenuity for the proper control of artifacts. The important outcome is that the method developed holds a promise of better streamlining in future cockpits.

Another of our familiar old problems arose with respect to aeroembolism and projected Apollo missions. The question was how to minimize the incidence and the severity of bends if ascent should require a stay of three hours or longer at a 35,000 foot altitude. Tests were run on 12 subjects in one of our low pressure chambers. The subjects represented a broad cross-section of body types among Naval personnel. A relatively small but very closely controlled series of ascents yielded the following three alternative recommendations: (1) that Apollo astronauts should pre-breathe 100% oxygen at ground level for 3 hours or longer before the launch or (2) that



Fig. 1. Subject in Variform Capsule



Fig. 2. Tying Shoe Laces

they breathe 50% oxygen at 18,000 feet for 18 hours or longer before ascent to the 35,000 foot altitude or (3) that 100% oxygen be used on the ground for at least 2 hours before take-off, followed by 50% oxygen for at least 12 hours at 18,000 feet. These are conservative sounding recommendations, as they are indeed intended to be. Our chamber results were very consistent and provide convincing evidence that our subjects were more susceptible to aeroembolism than most of the other groups reported in the literature. Within our small group there were relatively wide divergences in age, adiposity, flying experience, chamber experience, and joint-sensitiveness occasioned by prior injury. Nevertheless, they were a highly-motivated group, and the deviations in their subjective reactions under graded environmental test and dummy conditions in the altitude chamber were strikingly uniform. These circumstances argue for the validity of our experimental conclusions and for their probable relevance in the Apollo application. The bends-susceptibility of our men should provide an operational factor of safety for astronauts. At the same time, this factor of safety appears realistic and not unreasonably conservative. Furthermore, the recommended alternatives of preoxygenation constitute a sliding schedule from which the least interference with operational needs and possibilities may be selected.

A third problem in which we were involved was to check the likelihood that the combined stresses of launch acceleration, 100% oxygen in the ambient environment, and re-entry acceleration might cause atalectasis during projected Gemini missions. This required weeks and weeks of round-the-clock efforts and widely coordinated support in different establishments. The subjects were 8 Navy and Marine Corps astronaut selectees. Two were used as experimental controls and as stand-bys in case of emergency need, and 6 were scheduled for the following experimental sequence: (1) a 10-day period of isolation in our BOQ while undergoing intensive batteries of preparatory medical examinations and psychomotor performance testing. On the last three days, the subjects were fed only a specified powdered diet. (2) On successive days, each of the 6 men was treated as follows: preparation for EKG and respiration monitoring, followed by attirement in a pressure suit, encasement in a portable altitude chamber simulating a Gemini capsule, and 4 hours of 100% oxygen breathing. Pure oxygen was the only gas supplied for breathing from this time to the end of the test. During this stage, the subject was transported in a specially fitted van to the Naval Air Development Center at Johnsville. Medical monitoring was continued during the trip. (3) After attachment to the arm of the human centrifuge in AMAL, the capsule was evacuated to a 27,000 foot equivalent altitude. Our colleagues at AMAL monitored the subject's condition and he then underwent two pulses of acceleration at 7G levels, simulating the first and second stages of launch. Figure 3 represents the sequence of events. (4) He was then returned, still inside the capsule, to ACEL, by means of cranes, hoists, and the van. (5) The entry port of the capsule was attached to the door of our Bio-Astronautic Test Facility by a hermetic seal and pressures were equalized. The

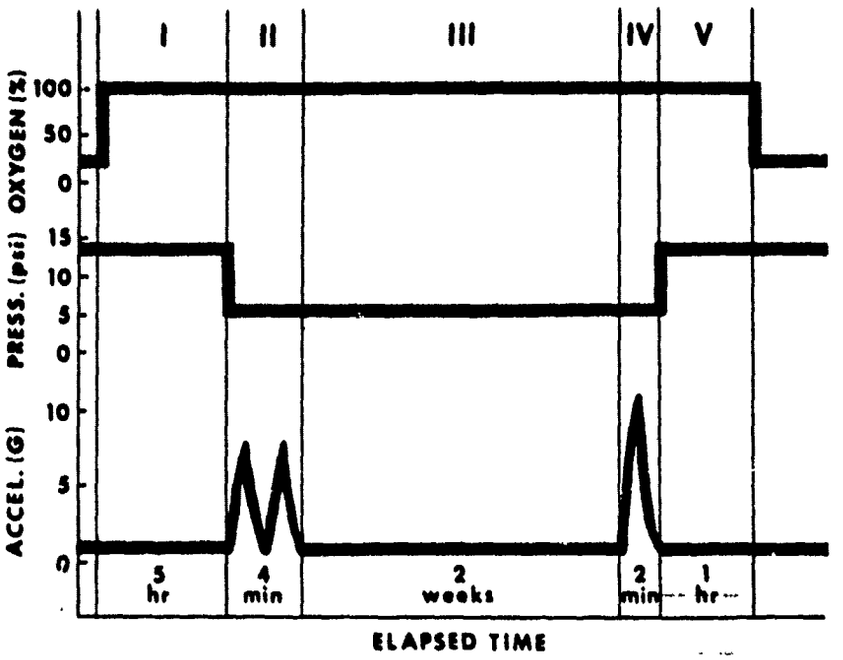


Fig. 3. Proposed GEMINI Profile

subject entered the chamber in which he was to live for the next 14 days and to simulate orbit life, and removed his pressure suit. (6)The pressure in the chamber was kept at 5 psia, and oxygen was bled in at a rate sufficient to keep the nitrogen concentration down. In addition to the daily routine of examinations and performance tests established previously in the BOQ, each subject now received chest x-rays inside the chamber ~~every third day~~. (7) ~~At the conclusion of the~~ 14-day exposures, the subjects were again instrumented for recording, re-packaged into the pressure suit and mobile capsule, and returned to the centrifuge as before. This time the stress was a simulated re-entry acceleration of 11G. (8) A three-day period of post-run testing concluded the schedule. Throughout the planning and conduct of these tests, we also received fine cooperation from specialists at the Philadelphia Naval Hospital and at the Hospital of the University of Pennsylvania.

Three of the men had been run through the entire experimental routine, when a chamber fire made it necessary to get the other three out in a hurry, and thus prevented further continuation. However, the results of the exposures to the combined stresses of oxygen and acceleration in these tests can be stated with considerable assurance. No untoward effects could be ascribed to the acceleration alone. During simulated orbit-life in the low pressure chamber, the only undesirable symptom observed by all six subjects was mild chest discomfort which developed after the first or second day, and then tended to subside. Neither the tests nor x-ray photographs during the run, nor subsequent tests of lung compliance by Dr. Arthur B. DuBois at the University Hospital indicated the occurrence of any organic or persistent functional pulmonary derangement. Slight reductions in peripheral visual fields and in dark adaptation seemed to develop in some of the subjects and in the medical officer who accompanied them during the chamber confinement. However, the quantitative assessment of the observations is incomplete because of inescapable differences in the visual test conditions inside and outside the chamber. The medical observer entered the chamber with mild nasopharyngitis. He soon had a violent bilateral serous otitis but did manage to control it and hold out for the entire duration of the run. Two of the subjects also developed serous otitis early during the run, but they recovered rapidly and without further incident. It has therefore, been concluded that the proposed Gemini acceleration profile and breathing schedule is not likely to cause atalectasis.

A fourth investigation which I should like to describe dealt with negative impact tolerance under conditions that had been proposed for manned Apollo flights. The problem was to determine whether seated subjects could be exposed safely to downward impacts at accelerations ranging up to 20G and at rates of acceleration increase up to 2500 G/sec. The general plan was to seat a so-called anthropomorphic dummy on its back on our horizontal accelerator, attach it with lap and shoulder harnesses, and then apply a mild impulsive force from the direction of the head, to propel the sled

and dummy away feet first. If high speed motion pictures and acceleration recordings offered no contra-indications, we were then to expose human volunteers to this same mild impact and then proceed slowly to progressively higher G loadings. Unfortunately, the results had been needed prior to the work assignment, which introduced an undesirable undercurrent of hastiness, in feeling, if not in fact. Also, it had been assumed by the requesting agency that exact details about body attitude, support and restraint, or about limitations to these conditions, need not be considered very precisely at the start. In working toward 20 negative G at 2500 G/sec, this assumption cannot be maintained. So far as human tolerance is concerned, this is a largely unknown and perilous zone in the G-versus-time function. The skeletal structures of the spine have been supposed to be most susceptible to injury, since longitudinal accelerations of 20 to over 40 G may be anticipated in the subject. But the effects of the transmitted and rebounding impact forces upon soft tissues within the head, chest and abdomen may conceivably cause still more serious injury than, say, a broken bone. You may recall an old account in the BUMED NEWS LETTER about some swing impact tests carried out in Germany by Henschke, in which 10 G transmitted to the base of the skull appeared to be the threshold for human concussion, to mention only one of the warning notices in the available literature.

Nevertheless, we proceeded rapidly. The tests were conducted by a very competent engineering team, and human exposures took place only under the close surveillance and direction of a flight surgeon. The criteria for evaluating tolerance included instrumented physical determinations and photography, biological recordings, medical examinations and radiographs before and after each run, as well as subjective reports. From the very start of the impact exposures, a succession of new hazards had to be skirted by experimental redesign. The system for transferring power from the ram to the sled had to be modified to maintain the desired shape of the acceleration pattern as loading was raised from the lower to the higher G levels. The subject's head had to be encased in a helmet and tied down to his torso to keep it from whipping up and down. The seat angle between the thighs and the back had to be changed from slightly obtuse, to rectangular or slightly acute. Harness had to be widened, placed on additional locations on the body, and attached from more effective points of restraint on the seat. Skillful instrumentation had to become even more highly sophisticated. At present we have reached a stage in the neighborhood of 15 G application, and are trying to gather together the threads for the remaining pull. The violent stroke of the ram is over in less than 100 milliseconds, and the recoil is brutal. The next figure 4, may give you some idea of how a subject appeared to be getting pulled apart and immediately squashed together in one of the tests. This one was with only 10 G at 1140 G/sec. In this particular run, we obtained an x-ray picture of the thoracolumbar spine at the time of peak G, something that is very difficult and perhaps unique in impact tests. Fortunately, we had very able cooperation from the Philadelphia Naval Hospital in

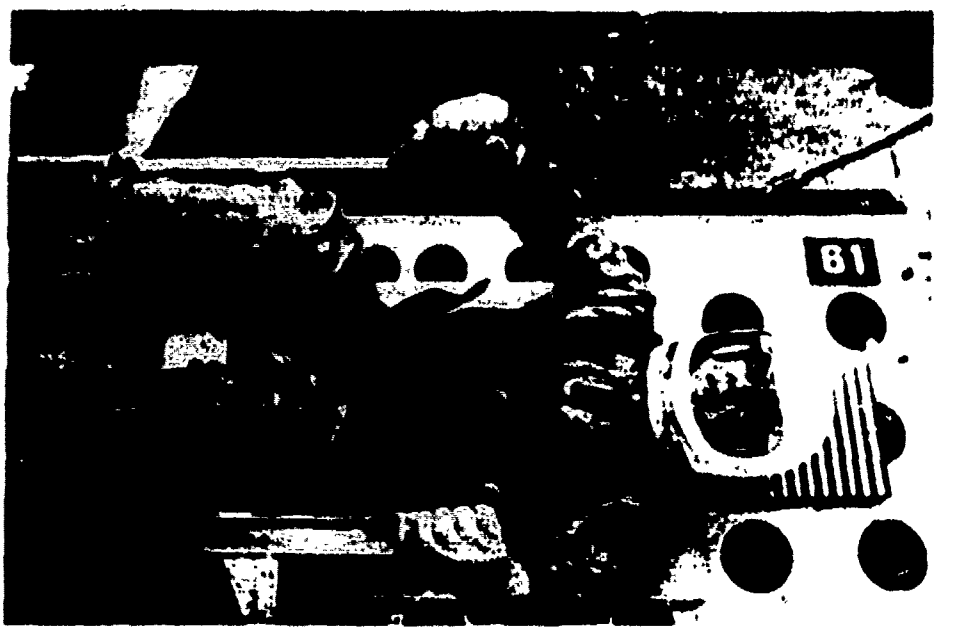


Fig. 4. Subject Undergoing 10G Impact at 1 140 G/Sec.

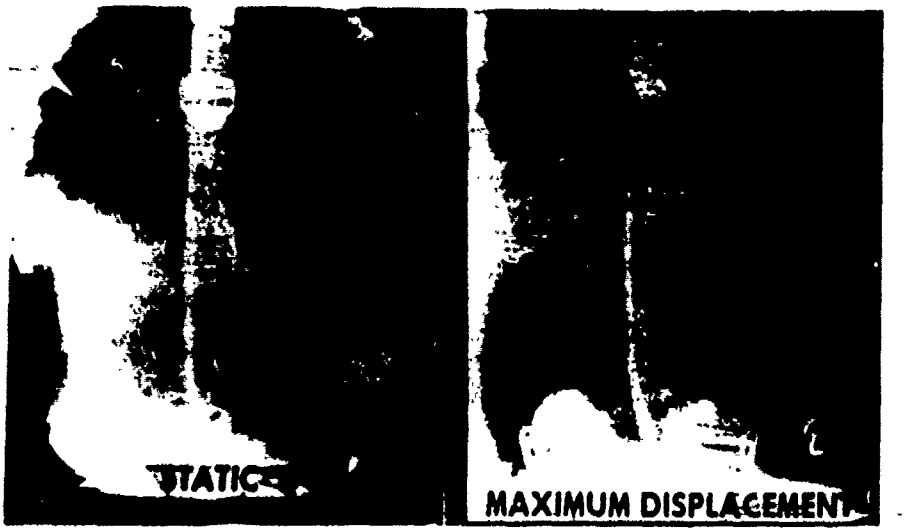


Fig. 5. X-ray of Thorocolumbar Spine,
10G Impact at 1140 G/Sec.

securing this result. It is shown in figure 5. You can see that we have certainly made progress. But the epitome of what we have accomplished in this task to date is that we have learned how to go about preparing to enter the danger zone ahead of us.

These four examples of our work have neglected much of the diversity among our assigned tasks. As you may know, our prime function is that of a hardware laboratory for the Bureau of Naval Weapons. There is nothing in the line of personal safety and protective equipment in naval aviation that does not enter our portals at some point. There are quite a few such items that were conceived, developed and perfected in our laboratory. Needless to say, very few of these contributions would have been possible if we had depended upon only our internal resources. We have carried on continual, many-sided collaborative exchanges of information and of services with other Federal and Defense Establishment activities, with industrial and technical specialists and organizations, and with the general scientific community. We therefore, regard ourselves less as mere customers for advanced technology than as one of the pioneering members of the Navy's own technological team.

ELASTIC AND PLASTIC BEHAVIOR IN THE FINITE DEFORMATION OF METALS BY IMPACT AND EXPLOSIVE LOADING

Marvin E. Backman
U. S. Naval Ordnance Test Station
China Lake, California

INTRODUCTION

Many problems connected with the operation and effectiveness of weapons require information on the behavior of metals when subjected to intense loading at extremely rapid rates. Two examples are (1) the cratering of target from the impact by a projectile, a fundamental problem in terminal ballistics, and (2) the response of a metal to the detonation of an explosive in contact with it, a fundamental problem in warhead technology. In both cases the stresses may be of the order of 100 kilobars or better and large strains are one of the most outstanding characteristics. The present paper describes an application of dynamic theories of elasticity and plasticity for determining material behavior under these conditions.

Consider a specific example. Figure 1 shows the section of a crater formed by the impact of cylindrical steel projectile against a thick aluminum alloy target at a velocity of 1,200 meters/sec. It is estimated that the stresses developed during the formation of this crater may have been as high as 200 kilobars. The sectioned crater has been polished and metallurgically etched so that the internal deformations are revealed by the distortions of the "rolling texture." It can be seen that material has deformed extensively especially near the crater. The square I for example represents the estimated location and shape of an element of the target plate before the impact. That same element is now located and shaped as is shown by the parallelogram II. This element has been flattened in one direction by 30% and elongated in another direction by 50%. It has also been rotated by 20° and its sides have been sheared through 25°.

Perhaps the most important feature of the crater shown in Fig. 1 is that there is no evidence of internal changes other than the large mechanical deformations. If there had been extensive melting, for example, the size of the grains would be changed and the orientation of the grains would be random with respect to the distorted



FIG. 1. The Crater Produced By The Impact Of A Flat-Ended Cylindrical Projectile Against A Thick Aluminum Alloy Plate.

boundary shape. Studies have been made in lead in which the energy of the impact was sufficient to cause the melting and these indications of melting were in fact observed. Thus, the internal structure shown in Fig. 1 suggests that the deformation of the material is the main physical condition needed to account for material behavior; however, any analysis of such states of deformation must take into account the finite sizes of the deformations and the large stresses. The present work is an attempt to describe material behavior with this objective.

The assumption is made that for impact or explosive loading it is possible to define quantitative measures of distinct elastic and plastic strains and that the analysis of these deformations is sufficient as a basis for a theory of material behavior. It is further proposed to carry out the analysis of the deformations by modifications of the methods used for small strains and at conventional rates of loading. The principle tasks in the modifications are to take into account the finite size of the strains and the extremely rapid development of the strains and the stresses.

Considerable work has been done on impact and explosive loading which does in fact take into account the finite nature of the deformations and which explicitely postulates a nonlinear relation between the stresses and the strains (Ref. 1–3), but these have been based on the conclusion that the stresses consist of a relatively small elastic part plus a larger hydrostatic part. The small elastic part is then treated by linear elasticity and the hydrostatic part by the nonlinear techniques of fluid mechanics. The present approach does not necessarily disagree with this conclusion but it is the intention to investigate how this, or an equivalent conclusion comes about from the explicit introduction of theories of plastic strain into an analysis of material behavior.

HIGH STRAIN RATES

We shall consider first the problem of the modification of conventional concepts of plasticity and elasticity for the description of material behavior under very rapid development of stresses and strains. It is possible to treat this problem independently of the problem of accounting for the effects of finite strains. In impact studies, for example, attention may be restricted to special systems and velocities of impact that produce plane strains small enough to be treated by linear methods but which correspond to stresses sufficient to cause plastic deformation. For most metals and alloys there is a regime of impact velocities in which this condition is fulfilled. In this regime the simplification of small strain theory holds; the elastic behavior can be treated by the linear elasticity, and the total strain is the sum of the elastic and the plastic strains. The principal difficulties are in finding the proper relation between the stress and the plastic strain.

For the case of small strains the measure of the amount of strain can be expressed in terms of the displacements of the body by the formula

$$e_{ij} = \frac{\partial w_i}{\partial x_j} + \frac{\partial w_j}{\partial x_i}$$

where w_j is the displacement and x_i are the space variables. The elastic and plastic strains can be distinguished on the basis of the reversibility of the process of deformation. Thus, if an increment of work done in a given deformation is

$$dW = \sigma_{ij} de_{ij}$$

then

$$dW = dW_e + dW_p$$

where dW_e is the part of the work that is reversible and dW_p is the part that can not be recovered. The elastic and plastic strains then correspond to these contributions to the total work (Ref. 4).

$$\sigma_{ij} de_{ij} = \sigma_{ij}(de^E_{ij} + de^p_{ij})$$

$$e_{ij} = e^E_{ij} + e^p_{ij}$$

If it is assumed that the work done by the plastic deformations is the only contribution to the heat term of the energy balance then the work that is recoverable is the same as a change in the elastic potential. Furthermore, this potential is not appreciably affected by the plastic deformation so that the elastic constants are the same as before deformation, thus

$$\sigma_{ij} = \lambda \sum_{i=1}^{i=3} e^E_{ii} + 2\mu e^E_{ij}$$

$$\sigma_{ij} = \lambda \sum_{i=1}^{i=3} e_{ii} + 2\mu(e_{ij} - e^p_{ij})$$

In the above the plastic deformation is associated in some way with the irreversible part of the deformation. To make this association of plastic strain and irreversibility more specific we may simply take it as an empirically observed fact or appeal to molecular

and crystallographic notions. In any case the plastic deformation ultimately must be related to the stress (or some other condition of the body) in a way that is completely distinct from the elastic relation.

In the present discussion we choose to relate the plastic strain rate to the stress and thus we follow the procedure first proposed by Malvern (Ref. 5). Many arguments can be presented both for and against this choice of variables. In most cases these arguments involve questions outside the interests of the present work and therefore this choice of variables is simply postulated, appropriate data are used to describe the plastic behavior, and the consequences of the choice are then developed. The final results can be compared to experimental results that have been reported recently and which tend to support the present choice.

As a mathematical representation of the relation between the plastic strain and the stress we choose

$$\dot{e}^{p}_{ij} = \phi S_{ij} \quad \text{where} \quad S_{ij} = \sigma_{ij} - \left\{ \frac{\sigma_{11} + \sigma_{22} + \sigma_{33}}{3} \right\}$$

The quantity ϕ serves to specify the properties of dynamic plasticity including the phenomenon on yielding. In place of a criterion for the discontinuous onset of plastic deformation the quantity ϕ contains essentially the same effect through very rapid but continuous changes in the vicinity of these critical combinations of stresses. An appropriate form for the quantity ϕ is

$$\phi = A \exp F(\sigma_{ij}, e^{p}_{ij})$$

The exponential function is a mathematically simple expression that can adequately describe the changes of strain rate that are required. The functions F are here considered to be given either by the results of separate theoretical analysis or as representations of empirical data. In the present discussion the latter use is made of these functions and for this reason the simple form is used

$$F = B\left[\sqrt{S^2_{11} + S^2_{22} + S^2_{33} + S^2_{12} + S^2_{13} + S^2_{23}} - \text{constant}\right]$$

to approximate selected intervals of experimental data obtained by Hauser, Simmons, and Dorn (Ref. 6). An example of these data is shown in Fig. 2 with the above form of approximation to it.

It is most convenient to combine this expression for the plastic strain rate directly with an expression for the stress rate derived from the elastic stress-strain relation above, which gives

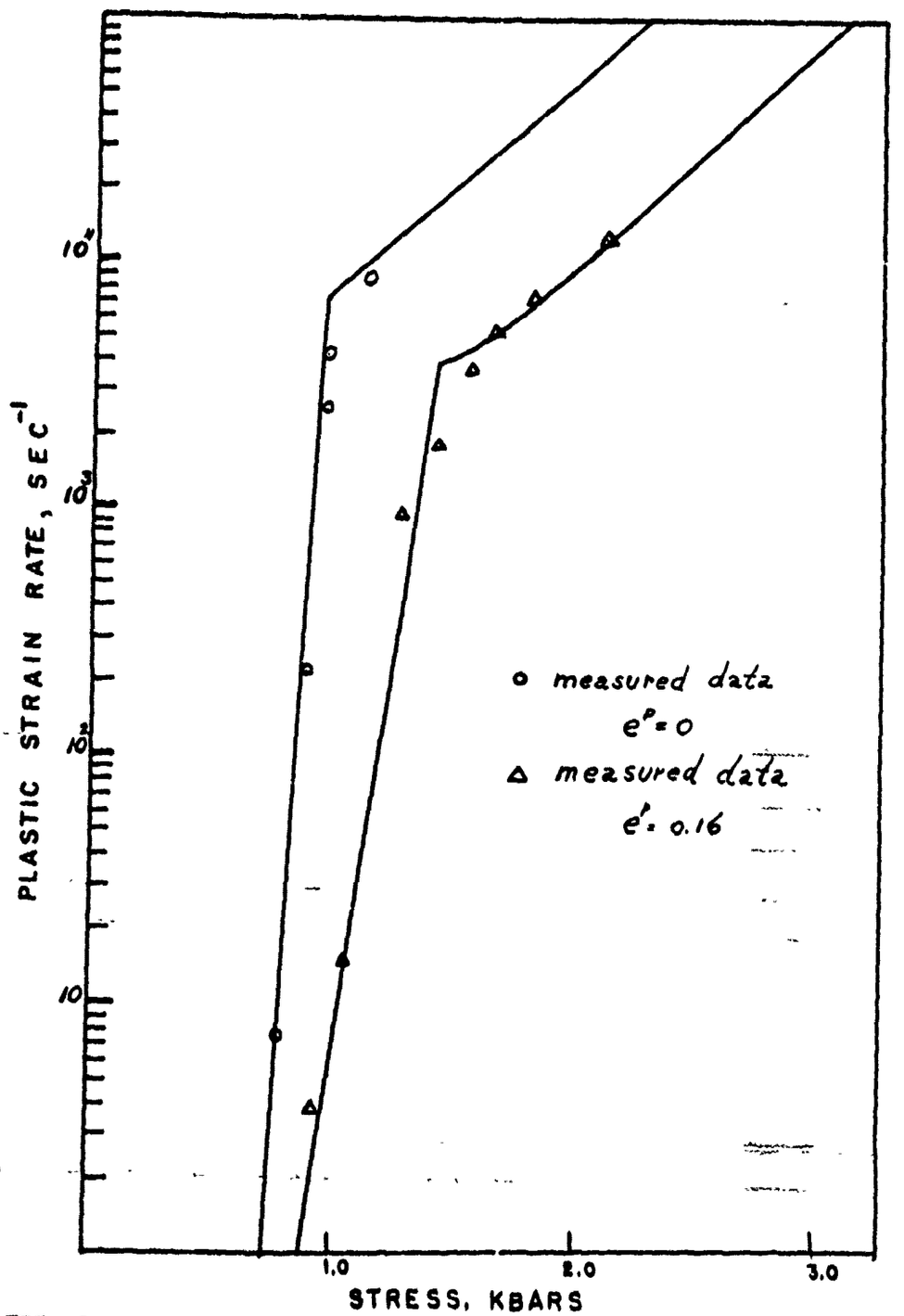


FIG. 2. A Graphical Representation Of The Mathematical Relation Between The Plastic Strain Rate And Stress And A Comparison To Experimental Data.

$$\frac{\partial \sigma_{ij}}{\partial t} = (\lambda + 2\mu) \frac{\partial}{\partial t} (e_{11} + e_{22} + e_{33}) + 2\mu\left(\frac{\partial e_{ij}}{\partial t} - \phi s_{ij}\right)$$

This equation can then be used as the equation of material behavior in the problem of determining plane wave propagation.

The equations governing one-dimensional plane wave propagation consist of the equation of continuity, the equation of motion, and the equation of material behavior. For the elastically linear problem considered here, these are

$$\frac{\partial e_1}{\partial t} - \frac{\partial v}{\partial x} = 0$$

$$\rho_o \frac{\partial v}{\partial t} + \frac{\partial \sigma}{\partial x} = 0$$

$$\frac{\partial e_1}{\partial t} - \frac{\partial \sigma_1}{\partial t}/(\lambda + 2\mu) = \left(\frac{2\mu}{\lambda+2\mu}\right) \phi \frac{2}{3}(\sigma_1 - \sigma_2)$$

where

$$\sigma_2 = \lambda e_1 + \mu e_1^p$$

These may be solved numerically by the method of characteristics. One form for presenting the solutions is to show the computed stress pulse shape at some fixed distance. Examples are shown in Fig. 3.

These results may be compared to experimental measurements of the pulse shapes measured after transmission through various metals and alloys (Ref. 7-9) in Fig. 4. The calculations and the measurements are in essential agreement in the slope of the pulse shape from the elastic rise to the "plastic" part of the pulse that follows. The elastic part of both the computed and the measured pulses decrease with the distance traveled through the material until some stable value is reached.

One important conclusion from this calculation is that the material behavior at the start of the process is essentially elastic and approached a state composed of an elastic and a hydrodynamic state by a continuous transition. If one is interested in the transition from the initially elastic to the final state it is necessary to use a method that is appropriate for the initial elastic state. It is obvious that one has this interest in problems involving thin targets or explosive liners. For the present special case where the strains are small, linear elasticity is adequate but when the stresses are large enough to give finite strains then the methods of nonlinear elasticity are required and with some appropriate means for introducing the plastic strain.

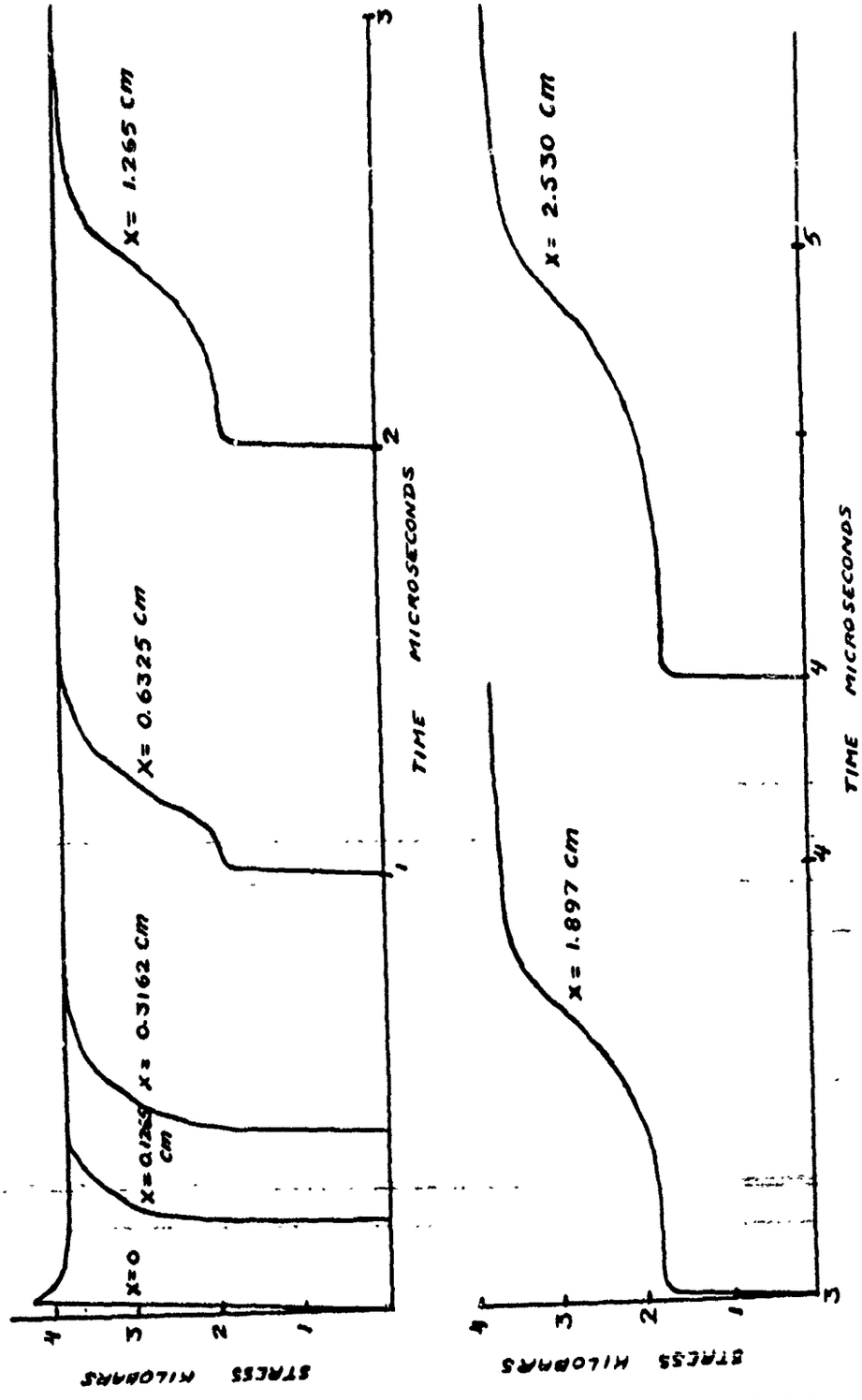


FIG. 3. A Graphical Representation Of The Computed Pulse Shapes At Various Distances Into An Aluminum Slab.

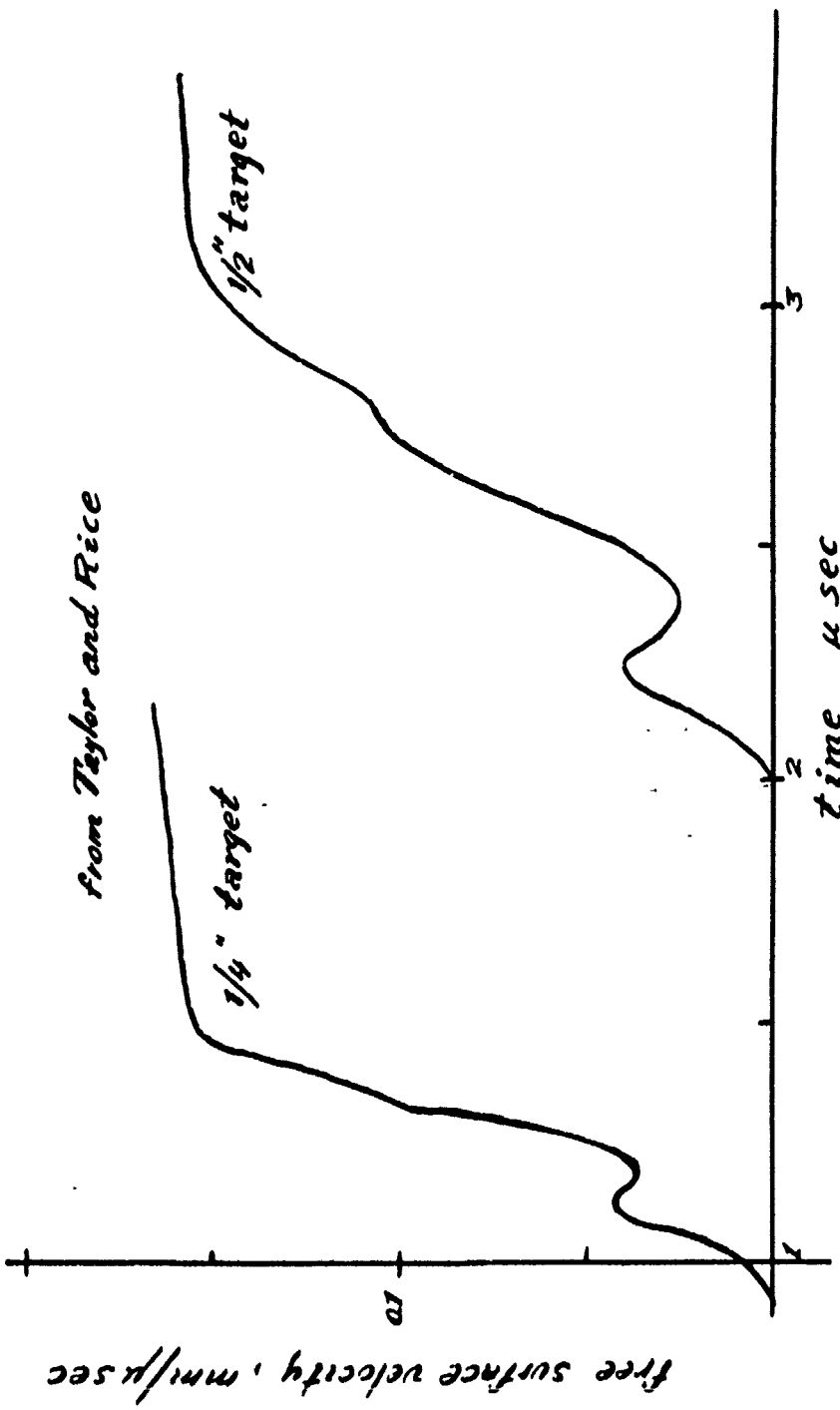


FIG. 4. Measurements Of The Pulse Shape At The Free Surface Of Two Slabs Of Different Thicknesses.

FINITE DEFORMATIONS

Finite deformations of a body are not accurately measured by the linear expression for the strain in terms of the displacements w_i of the body. An accurate form is given by (Ref. 10)

$$\epsilon_{ij} = \frac{1}{2}\left\{\frac{\partial w_i}{\partial x_j} + \frac{\partial w_j}{\partial x_i} - \frac{\partial w_k}{\partial x_i}\frac{\partial w_k}{\partial x_j}\right\}$$

In the problems of interest to the present discussion the strains of the body are the results of elastic and plastic contributions. Thus, the question arises how are these elastic and plastic contributions given a strain measure? In order to see what is involved in the answer to this question we consider a simple form of deformation in which there is no shearing of the material but only compression or extension of the material in the principal directions. One way of viewing the deformation process is to consider a fixed region in space and then observe how the material is deformed to fit into that region of space. In Fig. 5 this is geometrically represented for a single direction of straining. The square in Fig. 5a is the region fixed in space. The rectangle of Fig. 5b is the part of the body that was forced into this fixed region reversibly. The strain can be measured by the relative contraction of the body, i.e., the change in length divided by the final length

$$E^e = \frac{\Delta \ell}{\ell}$$

The deformation may also consist of an irreversible part. This means that at one time the part of the body now in the rectangle of Fig. 5b was previously in a larger rectangle shown in Fig. 5c. In order to define this deformation consistently with the elastic deformation the relative contraction of this deformation must be the change in length divided by the final length shown in Fig. 5b thus,

$$E^p = \frac{\Delta \ell'}{\ell'}$$

The total deformation is the result of the contraction of the large rectangle of Fig. 5c into the fixed region of Fig. 5a thus this deformation is given by

$$E = \frac{\Delta \ell''}{\ell}$$

These measures of the strain are not related as the small strain theory would predict. Rather the strains are related by

$$E = \frac{\Delta \ell''}{\ell} = \frac{\Delta \ell}{\ell} + \frac{\Delta \ell'}{\ell} = \frac{\Delta \ell}{\ell} + \frac{\ell'}{\ell}\frac{\Delta \ell'}{\ell'} = E^e + (1 - E^e)E^p$$
$$= E^e + E^p - E^e E^p$$

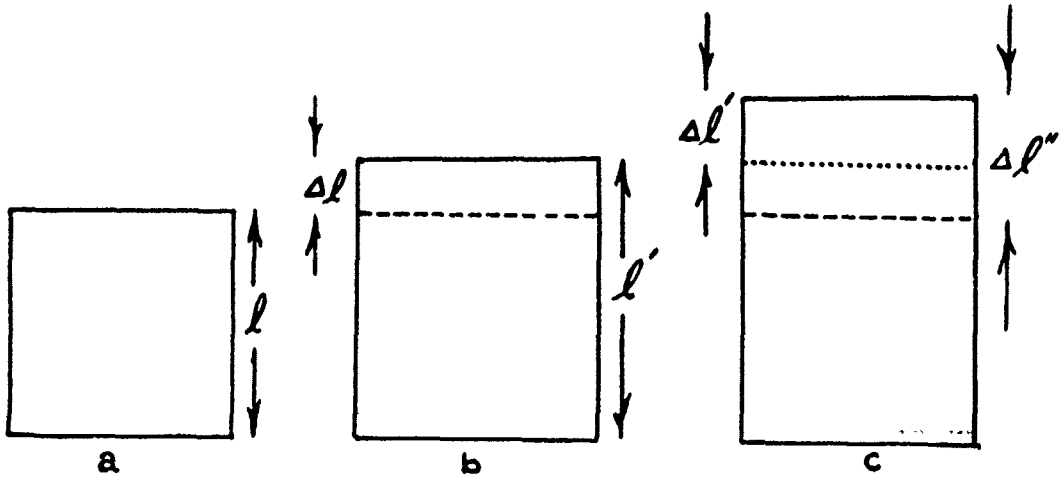


FIG. 5. A Diagram Of Elastic And Plastic Stages Of Deformation.

which reduces to the simple sum if the elastic strain is very small. This can be seen in Fig. 5 since the denominator of all the relative contractions is the same if the elastic strain is negligible and the sides of Fig. 5a and 5b are equal. This deviation from the linear relation between the strains is due to the necessity of measuring the plastic strain with respect to a different sized region than in the case of the elastic strain.

The above elementary ideas are a specific example of the general process of transforming one region of space into another region of space. In the entire process there are two transformations made between three regions. In the general mathematical representation the regions are given by three sets of labels for space coordinates. There is the fixed region of space with coordinates x_1, x_2, x_3, the intermediate region of space with coordinates y_1, y_2, y_3, which is transformed into the fixed region by the elastic process and the initial region with coordinates z_1, z_2, z_3, that is transformed into the final region of space by the total deformation. The strains can now be defined by the type of expression given earlier. The total deformation is given directly by that equation using the displacements from the initial to the final region. The elastic deformation is given by the same kind of equation using the displacements from the intermediate configuration to the fixed configuration of space.

$$\epsilon^E_{ij} = \frac{1}{2}\left\{\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{\partial u_k}{\partial x_i}\frac{\partial u_k}{\partial x_j}\right\}$$

The plastic deformation is given by the displacements from the initial configuration to the intermediate region of space and the partial differentiation is with respect to the coordinates of the intermediate region of space.

$$\epsilon^P_{ij} = \frac{1}{2}\left\{\frac{\partial v_i}{\partial y_j} + \frac{\partial v_j}{\partial y_i} - \frac{\partial v_k}{\partial y_i}\frac{\partial v_k}{\partial y_j}\right\}$$

In terms of the above definitions of strain one example of the relation between the total, elastic, and plastic strains is

$$\epsilon_{11} = \epsilon^E_{11} + \epsilon^P_{11} - 2\epsilon^E_{11}\epsilon^P_{11}$$

and the elastic strain component is given by

$$\epsilon^E_{11} = (\epsilon_{11} - \epsilon^P_{11})/(1 - 2\,\epsilon^P_{11})$$

Now it is possible to extend this kind of an argument to a much more general situation in which there are shears and rotations of the elements of the material. One case that is of particular use in

the solution of problems with impact or explosive loading is that of cylindrical symmetry. In this case the relation between the total strains, elastic strains, and plastic strains is given for the components as follows

$$\epsilon_{ij} = \epsilon^E_{ij} + \epsilon^p_{ij} - \epsilon^p_{il}\epsilon^E_{lj} - \epsilon^p_{il}\omega^E_{lj} - \epsilon^p_{jl}\epsilon^E_{li} - \epsilon^p_{jl}\omega^E_{li}$$

where

$$\omega^E_{ij} = \frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i}$$

These equations in turn can be solved to give the elastic strain components as a function of the total and elastic strain components for use in the stress-strain relations. One example is as follows

$$\epsilon^E_{11} = A(\epsilon^p_{ij})\left(\frac{\epsilon_{11} - \epsilon^p_{11}}{1 - 2\epsilon^p_{11}}\right) + B(\epsilon^p_{ij})\left(\frac{\epsilon_{22} - \epsilon^p_{22}}{1 - 2\epsilon^p_{22}}\right) + C(\epsilon^p_{ij})(\epsilon_{12} - \epsilon^p_{12})$$

The functions A, B, and C are polynomials in the quantities ϵ^p_{ij} and $(1 - 2\epsilon^p_{11})$. The most significant feature of the equation above is the presence of the difference terms for strain components other than ϵ_{11}. Thus, for this case of cylindrical symmetry the elastic strains depend not only on the total and plastic strain components but also on other components so that there are cross-effects in the expression for the elastic strains.

STRESS-STRAIN RELATIONS FOR FINITE DEFORMATIONS

In the treatment of perfect elasticity the stress is derived from an elastic potential. The elastic potential may be given a thermodynamic interpretation as either the specific free energy, for an isothermal process, or as the specific internal energy, for an adiabatic process. The adiabatic process is appropriate for impulsive loading.

The starting point, in perfect elasticity, is the first law of thermodynamics,

$$\rho\delta u = \rho T\delta S + \sigma_{ij}\frac{\partial \delta x_i}{\partial x_j}$$

In the above, the quantity $\sigma_{ij}\,\partial\delta x_i/\partial x_j$ is the work done by increments of displacements δx_i against stresses σ_{ij}, and the specific internal energy u is assumed to be a function of the strain and the entropy $u(\epsilon_{ij},S)$. From this it follows that

$$\delta u = \frac{\partial u}{\partial S}\delta S + \frac{\partial u}{\partial \epsilon_{ij}}\delta\epsilon_{ij}$$

We may use the relation

$$\delta\epsilon_{ij} = (\delta_{ik} - 2\epsilon_{ik}) \frac{\partial \delta x_j}{\partial x_k}$$

which connect the strain and the quantities $\partial \delta x_j / \partial x_k$, (see Ref. 1)

$$\delta u = \frac{\partial u}{\partial S} \delta S + \frac{\partial u}{\partial \epsilon_{ij}} (\delta_{ik} - 2\epsilon_{ik}) \frac{\partial \delta x_j}{\partial x_k}$$

and thus arrive at the value for the stress, for a reversible process,

$$\sigma_{ij} = \rho(\delta_{ik} - 2\epsilon_{ik}) \frac{\partial u}{\partial \epsilon_{jk}}$$

One task of the present work is to modify the arguments above so that permanent deformations can be included. This task is accomplished by showing that it is possible to resolve the work term of the first law into two parts, one that contributes to specific internal energy and one that contributes to the heat term.

The points of the final configuration may be given as functions of the intermediate configuration,

$$x_i = x_i(y_1, y_2, y_3, t)$$

A variation of the quantity is now given by

$$\delta x_i = \frac{\partial x_i}{\partial t} \delta t + \frac{\partial x_i}{\partial y_j} \frac{\partial y_j}{\partial t} \delta t$$

$$= \delta x_i^E + \delta x_i^P$$

Since the quantities y_1, y_2, y_3, are assumed to vary with time, the variation has two parts. One part comes from the direct dependence of the final configuration on time, and is the elastic part of the variation since it is independent of the intermediate configuration. The other part of the variation comes from the dependence on the quantities y_1, y_2, y_3, and the dependence of these on time. This part of the variation is the plastic part since it is explicitly dependent on the intermediate configuration.

The work term of the expression for the first law becomes

$$\sigma_{ij} \frac{\partial \delta x_i}{\partial x_j} = \sigma_{ij} \left(\frac{\partial \delta x_i^E}{\partial x_j}\right) + \sigma_{ij} \left(\frac{\partial \delta x_i^P}{\partial x_j}\right)$$

Now the assumption is made that the work $\sigma_{ij}\, \partial \delta x_i^P / \partial x_j$ is dissipated as heat so that

$$\sigma_{ij} \frac{\partial \delta x_i^P}{\partial x_j} = \rho T \delta S$$

If this is the only irreversible process then the first law is

$$\rho \delta u = \rho T \delta S + \sigma_{ij} \frac{\partial (\delta x_i^E)}{\partial x_j} + \sigma_{ij} \frac{\partial (\delta x_i^P)}{\partial x_j}$$

$$= \sigma_{ij} \frac{\partial \delta x_i^E}{\partial x_j}$$

and consequently the stress is given by

$$\sigma_{ij} = \rho (\delta_{ik} - 2\epsilon_{ik}^E) \frac{\partial u}{\partial \epsilon_{jk}^E}$$

STRESS-STRAIN RELATIONS FOR AN ISOTROPIC SOLID

It is convenient to restrict attention to an isotropic material. The material properties are easily expressed and the description is reasonably accurate for many metals and alloys.

For an isotropic material the specific internal energy is a function of the strain through certain invariants of the strain

$$I_1^E = \epsilon_{11}^E + \epsilon_{22}^E + \epsilon_{33}^E$$

$$I_2^E = \begin{vmatrix} \epsilon_{11}^E & \epsilon_{12}^E \\ \epsilon_{12}^E & \epsilon_{22}^E \end{vmatrix} + \begin{vmatrix} \epsilon_{22}^E & \epsilon_{23}^E \\ \epsilon_{23}^E & \epsilon_{33}^E \end{vmatrix} + \begin{vmatrix} \epsilon_{33}^E & \epsilon_{13}^E \\ \epsilon_{13}^E & \epsilon_{11}^E \end{vmatrix}$$

$$I_3^E = \begin{vmatrix} \epsilon_{11}^E & \epsilon_{12}^E & \epsilon_{13}^E \\ \epsilon_{12}^E & \epsilon_{22}^E & \epsilon_{23}^E \\ \epsilon_{13}^E & \epsilon_{23}^E & \epsilon_{33}^E \end{vmatrix}$$

The specific internal energy may be given as a series

$$\rho_o u = \alpha_1 (I_1^E)^2 + \alpha_2 I_2^E + \alpha_3 (I_1^E)^3 + \alpha_4 I_1^E I_2^E + \alpha_5 I_3^E \cdots$$

and the density is

$$\rho = \rho_o \sqrt{1 - 2I_1 + 4I_2 - 8I_3}$$

hence the stress becomes

$$\sigma_{ij} = a_1 I_1^E \delta_{ij} + a_2 \epsilon_{ij}^E + a_3 (I_1^E)^2 \delta_{ij} + a_4 I_1^E \epsilon_{ij}^E + a_5 I_2^E \delta_{ij}$$
$$+ a_6 co\epsilon_{ij}^E + 2a_2 \epsilon_{ik}^E \epsilon_{kj}^E$$

The expression above can be given in terms of the total and the plastic strain components. The stresses are then

$$\sigma_{11} = a_1 I_1^E + a_2 \left(\frac{\epsilon_{11} - \epsilon_{11}^p}{1 - 2\epsilon_{11}^p}\right) + a_3 (I_1^E)^2 + a_4 I_1^E \left(\frac{\epsilon_{11} - \epsilon_{11}^p}{1 - 2\epsilon_{11}^p}\right)$$
$$+ a_5 I_2^E + a_6 \left(\frac{\epsilon_{22} - \epsilon_{22}^p}{1 - 2\epsilon_{22}^p}\right)\left(\frac{\epsilon_{33} - \epsilon_{33}^p}{1 - 2\epsilon_{33}^p}\right) + 2a_2 \left(\frac{\epsilon_{11} - \epsilon_{11}^p}{1 - 2\epsilon_{11}^p}\right)^2$$

$$\sigma_{22} = a_1 I_1^E + a_2 \left(\frac{\epsilon_{22} - \epsilon_{22}^p}{1 - 2\epsilon_{22}^p}\right) + a_3 (I_1^E)^2 + a_4 I_1^E \left(\frac{\epsilon_{22} - \epsilon_{22}^p}{1 - 2\epsilon_{22}^p}\right)$$
$$+ a_5 I_2 + a_6 \left(\frac{\epsilon_{11} - \epsilon_{11}^p}{1 - 2\epsilon_{11}^p}\right)\left(\frac{\epsilon_{33} - \epsilon_{33}^p}{1 - 2\epsilon_{33}^p}\right) + 2a_2 \left(\frac{\epsilon_{22} - \epsilon_{22}^p}{1 - 2\epsilon_{22}^p}\right)^2$$

$$\sigma_{33} = a_1 I_1^E + a_2\left(\frac{\epsilon_{33} - \epsilon_{33}^p}{1 - 2\epsilon_{33}^p}\right) + a_3 (I_1^E)^2 + a_4 I_1^E \left(\frac{\epsilon_{33} - \epsilon_{33}^p}{1 - 2\epsilon_{33}^p}\right)$$

$$+ a_5 I_2 + a_6\left(\frac{\epsilon_{11} - \epsilon_{11}^p}{1 - 2\epsilon_{11}^p}\right)\left(\frac{\epsilon_{22} - \epsilon_{22}^p}{1 - 2\epsilon_{22}^p}\right) + 2a_2\left(\frac{\epsilon_{33} - \epsilon_{33}^p}{1 - 2\epsilon_{33}^p}\right)^2$$

$$I_1^E = \left(\frac{\epsilon_{11} - \epsilon_{11}^p}{1 - 2\epsilon_{11}^p}\right) + \left(\frac{\epsilon_{22} - \epsilon_{22}^p}{1 - 2\epsilon_{22}^p}\right) + \left(\frac{\epsilon_{33} - \epsilon_{33}^p}{1 - 2\epsilon_{33}^p}\right)$$

$$I_2^E = \left(\frac{\epsilon_{22} - \epsilon_{22}^p}{1 - 2\epsilon_{22}^p}\right)\left(\frac{\epsilon_{33} - \epsilon_{33}^p}{1 - 2\epsilon_{33}^p}\right) + \left(\frac{\epsilon_{33} - \epsilon_{33}^p}{1 - 2\epsilon_{33}^p}\right)\left(\frac{\epsilon_{11} - \epsilon_{11}^p}{1 - 2\epsilon_{11}^p}\right) + \left(\frac{\epsilon_{11} - \epsilon_{11}^p}{1 - 2\epsilon_{11}^p}\right)\left(\frac{\epsilon_{22} - \epsilon_{22}^p}{1 - 2\epsilon_{22}^p}\right)$$

This stress-strain relation can be used to obtain a nonlinear form of constitutive relation for use in the solution of planar one-dimensional stress wave propagation. As in the linear problem it is convenient to use a differential form of the equation. We simply differentiate the stress-strain relations to obtain the local rate of change of stress and propose to use this with appropriate convective terms to represent the total rates of change for an Eulerian representation of the problem. The partial derivatives with respect to time are

$$\frac{\partial \sigma_{11}}{\partial t} = \left\{a_1 + a_2 + a_3(\epsilon_{11} - \epsilon_{11}^p) + \ldots\right\} \frac{\dot{\epsilon}_{11}}{(1 - 2\epsilon_{11}^p)} + \left\{a_2 + a_4(\epsilon_{11} - \epsilon_{11}^p) + \ldots\right\} \frac{\dot{\epsilon}_{11}^p}{(1 - 2\epsilon_{11}^p)}$$

where

$$a_1 = \lambda \quad \text{and} \quad a_2 = 2\mu$$

(The brackets above contain other terms beyond those shown but these are omitted for simplicity of writing.) The quantities $\lambda + 2\mu$ and 2μ of the linear case are replaced by

$$\lambda + 2\mu \rightarrow (\lambda + 2\mu + a_3(\epsilon_{11} - \epsilon_{11}^p) + \ldots)/(1 - 2\epsilon_{11}^p)$$

$$2\mu \rightarrow (2\mu + a_4(\epsilon_{11} - \epsilon_{11}^p) + \ldots)/(1 - 2\epsilon_{11}^p)$$

and the nonhomogeneous term is given by

$$\frac{\left\{2\mu + a_4(\epsilon_{11} - \epsilon_{11}^p) + \cdots\right\} \quad \phi\left\{\frac{2}{3}(\sigma_1 - \sigma_2)\right\}}{\left\{\lambda + 2\mu + a_3(\epsilon_{11} - \epsilon_{11}^p) + \cdots\right\}}$$

This form of the constitutive equation has the main features of the ~~constitutive equation used for the linear problem~~ but also accounts for the finite definition of strain, nonlinear stress-strain relations and the appropriate nonlinear relation between the total strain, elastic strain, and plastic strain. These last properties permit this type of constitutive relation to apply to situations in which it is expected that the propagation rate is dependent on the intensity of the disturbance.

The problem of one-dimensional wave propagation is important because of the simplicity both in experiment and theory; however, the urgent problems of the weapon designer are two- or three-dimensional. Some work is being done at other laboratories on two- or three-dimensional problems using linear elasticity and direct finite difference approximations to the basic equations. This same kind of work is contemplated at the Naval Ordnance Test Station using the constitutive relation appropriate for the case of cylindrical symmetry; however, there is no intention of attempting this kind of work until one-dimensional cases are better understood. There seems to be little point in trying to introduce all the complexity of the cross strain effects of the cylindrical case if the effect of the main strain component is still inaccurately known.

REFERENCES

1. Corenson, R. W., et al. "Dynamic Determinations of the Compressibilities of Metals," J APPL PHYS, Vol. 26 (1955), p. 1472-79.

2. Rice, M. H., R. G. McQueen, and J. M. Marsh. "Compression of Solids by Strong Shock Waves," in Solid State Physics, Vol. 6. New York, Academic Press, 1958. Pp. 1-63.

3. Morland, L. W. "The Propagation of Plane Irrotational Waves Through an Elastic-Plastic Medium," ROY SOC LONDON, PHIL TRANS, A251, No. 997 (1959), pp. 341-83.

4. Hill, R. The Mathematical Theory of Plasticity. London, Oxford Univ. Press, 1950.

5. Malvern, L. E. "The Propagation of Longitudinal Waves of Plastic Deformation in a Bar of Material Exhibiting a Strain-Rate Effect," J APPL MECH, Vol. 18, No. 2 (1951), pp. 203-8.

6. Hauser, F. E., J. A. Simmons, and J. E. Dorn. "Strain Rate Effects in Plastic Wave Propagation," in Response of Metals to High Velocity Deformation. New York, Interscience, 1961. Pp. 93-110.

7. Hughes, D. S., L. E. Gourley, and Mary F. Gourley. "Shock Compression of Bismuth and Iron," J APPL PHYS, Vol. 32, No. 2 (April 1961), pp. 624-29.

8. Jones, O. E., F. W. Nielsen, and W. B. Benedick. "Dynamic Behavior of Explosively Loaded Metals Determined by a Quartz-Transducer Technique," J APPL PHYS, Vol. 33, No. 11 (1962), pp. 3224-32.

9. Taylor, J. W., and M. H. Rice. "Elastic-Plastic Properties of Iron," J APPL PHYS, Vol. 34, No. 2 (1963), pp. 364-71.

10. Green, A. E., and J. E. Adkins. Large Elastic Deformations. London, Oxford Univ. Press, 1960.

CALOROBIC INORGANIC-ORGANIC POLYMERS

R. J. Landry
E. H. Bartel

U. S. Naval Ordnance Test Station
China Lake, California

During the past year, effort was directed primarily toward optimizing the preparation procedures and end-use of the zirconium (IV)phenoxyaldehyde polymer. A novolac (prepolymer) of hafnium(IV) phenoxyaldehyde and impure vanadium(V)phenoxyaldehyde was prepared. Aldehydes other than paraformaldehyde, and aromatic alcohols other than phenol have been evaluated. A continuing search for a better paraformaldehyde solvent is underway. Some progress has been made on a study, the ultimate goal of which is to obtain, as a pyrolytic product, the carbide composition 4TaC + HfC which has one of the highest melting points (7619°F) of any known material.

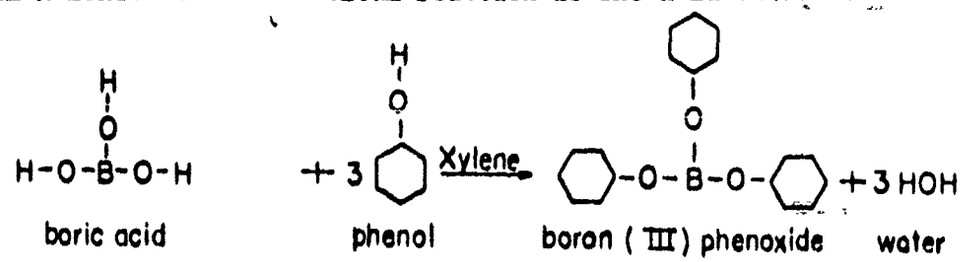
MAGNESIUM(II)PHENOXYALDEHYDE

Magnesium(II)phenoxyaldehyde has been prepared and test-fired as a micro-rocket-test-motor nozzle insert for 1/2-second. On test-firing for 3-seconds, a magnesium(II)phenoxyaldehyde-Refrasil insert eroded 0.003". The original throat diameter was 0.262". Longer firings would be required to establish the rate of erosion.

BORON(III)PHENOXYALDEHYDE RESIN

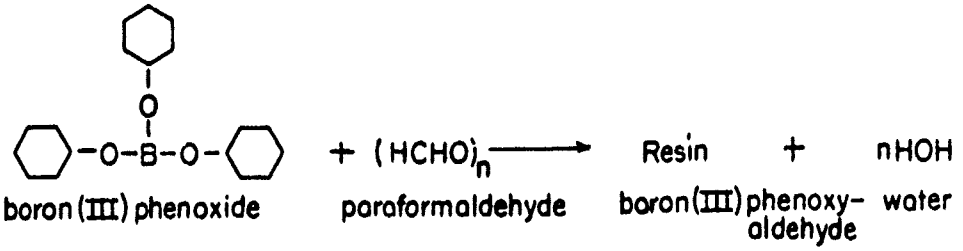
Preparation of boron(III)phenoxide and its resin was accomplished in one operation instead of the usual two.

Boron phenoxide was prepared by refluxing boric acid with phenol in xylene. The theoretical reaction is shown in Slide 1:



50 grams of boric acid was placed in a 50 ml flask with 268 grams of phenol and 150 ml of xylene, then refluxed for six hours. 33 ml of water was recovered. A small portion of the liquor was removed for evaluation. The white crystalline precipitate obtained from it was dried two hours at 122°F, removing most of the xylene, then its melting point (329°F) was determined.

A yellow resin was prepared from the clear liquid remaining in the flask by adding paraformaldehyde to it in small amounts, then refluxing for three hours. Slide 2 shows the theoretical reaction:



After adding half of the stoichiometric quantity (50g) of paraformaldehyde and refluxing for two hours, a clear, viscous liquid was obtained. The remaining paraformaldehyde was then added, and the material refluxed one more hour. 31 ml of water was recovered during resin formation. The viscous yellow polymer obtained after three hours refluxing, during which some xylene was lost, solidified on cooling. The polymer readily dissolved in tetrahydrofuran to give a clear solution.

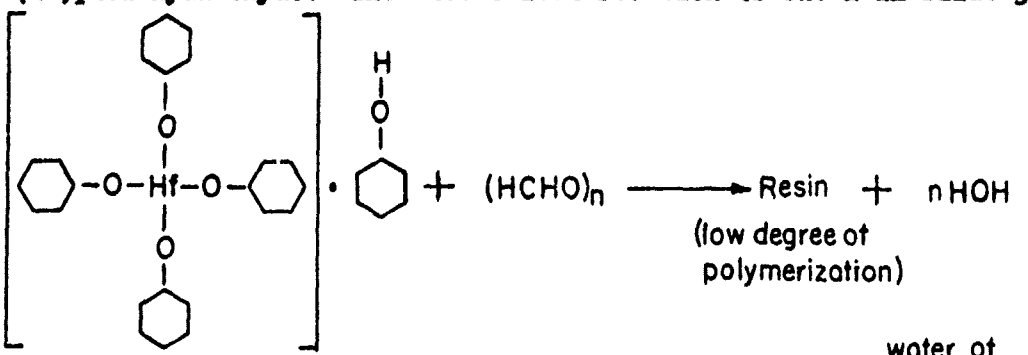
A laminate was prepared by dissolving some of the polymer in tetrahydrofuran and using the solution to impregnate glass cloth. The laminate was molded at 400°F for two hours under 10,000 psi pressure. It appeared to be a good laminate and had a Barcol hardness of 60-75.

The completeness of the reaction of the phenoxide and the resole polymer was determined in the following manner: In the phenoxide reaction a total of 33 grams of water was obtained, while the calculated stoichiometric amount was 44 grams. Some of the water may, of course, have been dissolved in the residue and the xylene. In the polymer formation 30 grams of water were obtained where the stoichiometric amount was 60 grams for complete polymerization to the resite stage. The polymer was in the resole stage, however, so only a fraction of the water of condensation was formed.

Assuming the paraformaldehyde is converted to methylene bridges (CH_2) and methylol end groups (CH_2OH), one may calculate the number of CH_2 groups converted (1:1 ratio with HCHO) then the remainder of the chemically combined water would be methylol groups. This procedure would help clarify the chemical structure of the polymeric end-product. It is assumed, of course, that no formaldehyde is lost in the polymerization reaction.

HAFNIUM(IV)PHENOXYALDEHYDE NOVOLAC

Previous attempts to prepare a novolac or prepolymer with hafnium have not been too successful, with the exception of that of hafnium (IV)phenoxyaldehyde: The theoretical reaction is shown in Slide 3:

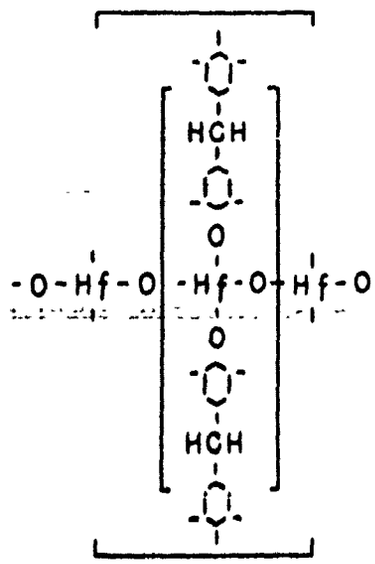


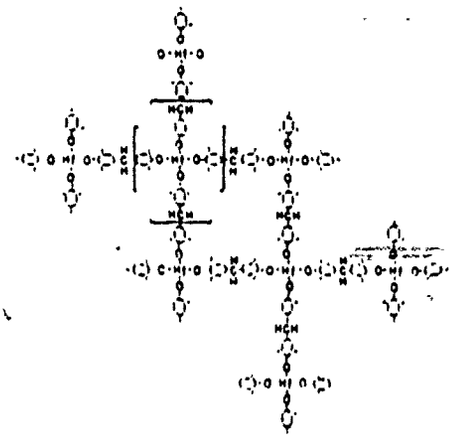
hafnium (IV) phenoxide (solvated) — paraformaldehyde — water of condensation

To prevent reactions of the hafnium(IV)chloride with the atmosphere, all transfers were performed in a dry box. Two hundred grams, or 0.625 moles of hafnium(IV)chloride was transferred into a one-liter three-neck flask and immediately covered with 100 ml of pure, dry benzene. The flask was fitted with a mechanical stirrer, a dropping funnel, and a nitrogen inlet. The dry nitrogen was used to flush the apparatus continuously during the preparation of the prepolymer. Two hundred ninety four grams of pure phenol was added a drop at a time. A vigorous exothermic reaction occurred. During the addition of phenol, the reaction mixture became a solid mass and dilution with more benzene (100 ml) became necessary. After adding the phenol, the reaction mixture was heated to refluxing temperature and refluxing was continued for two hours. After refluxing the product appeared as a red, transparent liquid. After cooling in an ice bath for 30 minutes, no precipitate appeared--nor did standing overnight produce a precipitate. A calculated amount (135 grams) of paraformaldehyde added in 27 gram aliquot portions, caused copious foaming. Heating to reflux temperature caused slight precipitation, which was filtered out. The filtrate was a viscous liquid containing 22% by weight of elementary hafnium .

Slide 4 shows the crystalline structure of the hafnium(IV)-phenoxide prepared by the resole method whereas by the described novolac method no crystals were formed. Slides 5 and 6 show two possible chemical structures of the hafnium(IV)phenoxyaldehyde resin.



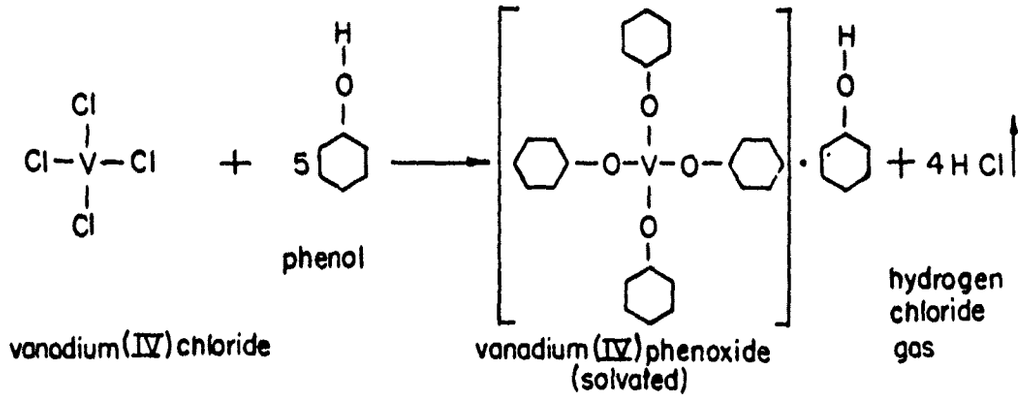




VANADIUM(IV)PHENOXYALDEHYDE

The reaction of vanadium trichloride with phenol does not proceed even after refluxing for 6 hours, with 5 times the stoichiometric quantity of phenol.

Vanadium(IV)phenoxide was then prepared from vanadium(IV)chloride. Slide 7 shows the theoretical reaction:



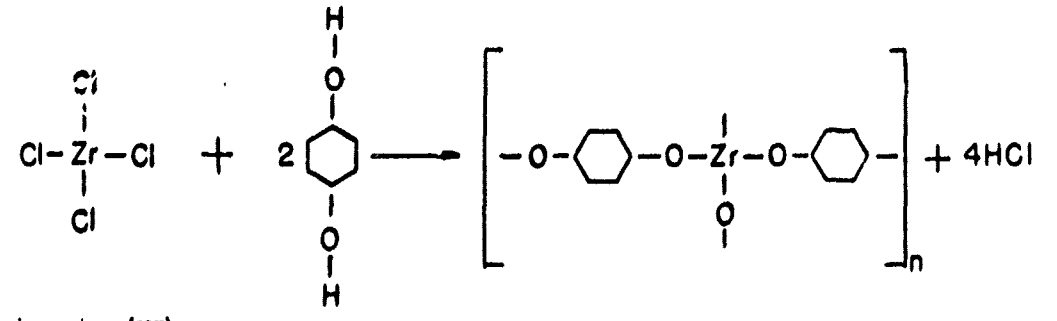
When 106cc of vanadium(IV)tetrachloride was added to 941 grams of phenol, then stirred and refluxed for 6 hours, hydrogen chloride was evolved. Benzene (500cc) was added and the solution was refluxed further, then cooled and filtered. The dark brown precipitate was washed 5 times with benzene, but remained somewhat mushy rather than becoming dry and crystalline in the manner of other phenoxides. The dark brown precipitate decomposes above 572°F and turns black when exposed to atmospheric moisture. Ignition of the phenoxide resulted in a yellow-red vanadium(V)pentaoxide, V_2O_5 The yield was 17.7% as compared with the theoretical quantity of 17.6% from the empirical solvated vanadiumphenoxide $(V(OC_6H_5)_4 \cdot C_6H_5OH$.

A black resin was prepared by mixing stoichiometric quantities of the phenoxide with paraformaldehyde, then heating the mixture in an oil bath. Tetrahydrofuran solvent was added when the resin temperature reached 212°F and the viscosity had increased sharply.

A laminate prepared with Refrasil had the low Barcol hardness of 10. Further tests were postponed until a purer batch of the phenoxide could be prepared.

ZIRCONIUM(IV)HYDROQUINIDE POLYMER or POLY[1,4,OXYBENZENE-ZIRCONIUM(IV)]

Zirconium(IV)chloride was reacted with hydroquinone. The theoretical reaction is as follows and is shown in Slide 8:



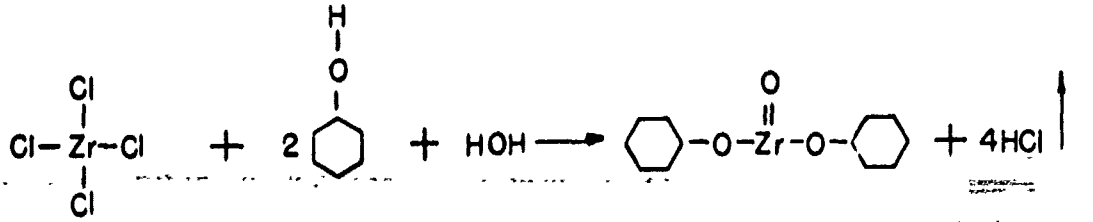
zirconium(IV) chloride — hydroquinone — polyzirconium(IV) hydroquinide

When stoichiometric quantities in accordance with this equation are mixed and heated, a rapid reaction occurs. At 400°F, a semi-solid polymer forms in one minute and hardens upon cooling. The viscosity increases rapidly and considerable foaming results from the evolution of hydrogen chloride.

This polymer could be powdered and, perhaps, cross-linked further with paraformaldehyde.

ZIRCONYL(IV)PHENOXYALDEHYDE

An attempt to prepare zirconyl(IV)phenoxide from zirconyl(IV)-chloride and excess phenol was unsuccessful. Another evaluation technique was to add two moles of phenol to one mole of $ZrCl_4$ in benzene, then boil it in an open beaker to allow atmospheric moisture to react with the chloride not replaced by phenol. The theoretical reaction without complex formation is shown in Slide 9:

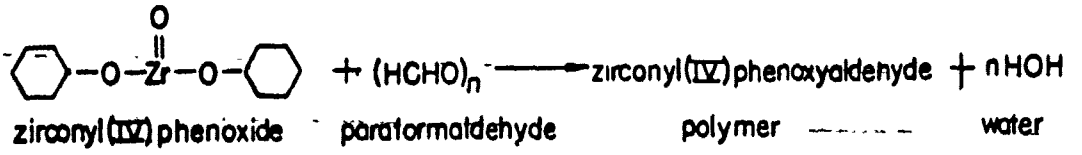


zirconium(IV) chloride — phenol — water — zirconyl(IV) phenoxide — hydrogen chloride gas

The reaction between the zirconium chloride and phenol proceeded at room temperature but, when the solution was heated to drive off the hydrogen chloride gas, the water reaction apparently produced a precipitate. On further heating to 284°F to drive off the benzene, white fumes were evolved and the odor of coal tar indicated decomposition.

A batch of zirconyl(IV)phenoxide was difficult to filter properly, so pentane and cyclohexane were added to the unpurified precipitate and the mixture then boiled in an open beaker on a hot plate. The pentane and cyclohexane boiled off and, at a gas temperature of 347°F to 365°F, a strong odor of phenol was noted. Boiling stopped and the temperature began to rise, so the reaction was interrupted and the mass allowed to cool and solidify. Analysis of the mass by ashing to ZrO_2 gave a Zr content of 27.6%. This is much more Zr than is contained in zirconium(IV)phenoxide, and is closer to the 30.88% metal content in zirconyl(IV)phenoxide.

The resin was prepared by adding a stoichiometric amount of paraformaldehyde to a dioxane solution of the zirconyl(IV)phenoxide described above, then heating from room temperature to boiling over a period of 6 minutes. The oil-bath temperature was 230°F. Slide 10 shows the theoretical reaction:



With 111g of zirconyl(IV)phenoxide, 36g of paraformaldehyde, and 85g of dioxane, the resin formed and began boiling at 212°F. It boiled excessively, so another batch was prepared in which the oil bath temperature was initially 149°F. The resin was slow to form, so after 30 minutes the temperature was increased to 176°F and held there 20 minutes. The resin temperature also reached 176°F, but did not exceed the oil-bath temperature. It was then cooled to 104°F in an ice bath. The resulting brown viscous resin contained 66% solids and only a very small amount of unreacted paraformaldehyde.

A laminate was made from Style 181 glass cloth impregnated with a 40% solution of the above-described resin in tetrahydrofuran. The cloth was dried 6 hours, then cut to make a 13-ply, 1/8"-thick laminate in a press at 325°F for 1/2 hour, under a pressure of 3000 psi. The resin flowed 1/16" to 1/8" beyond the cloth edges, hence, 10% by weight of the resin either flowed or was volatilized.

Two tensile specimens were machined in accordance with ASTM D638 Type III then tested to failure. The tensile strength was 38,000 psi. This was lower than 91LD phenolic resin Style 181 glass-cloth laminate that had been made up in a similar manner as a control sample.

ZIRCONIUM(IV)PHENOXIDE

Development of the zirconium(IV)phenoxyaldehyde resin and its evaluation was continued because it forms zirconium oxide as a pyrolytic by-product in a rocket-nozzle blast-exposed surface, and does not form the silicate as does titanium(IV)phenoxyaldehyde.

The theoretical zirconium content of the resin is uncertain since it is based on the combination of the stoichiometric quantities of the phenoxide and paraformaldehyde without allowance being made for loss of formaldehyde gas and water of condensation--factors which could easily result in the calculated zirconium content being too high. Hafnium is a natural impurity of zirconium, and is so similar to it, chemically, that its exact quantity is not easily established. A decrease in chlorine content during resin preparation is also expected.

The presence of chlorine is undesirable in the resin so a preparation of the phenoxide from zirconium(IV)acetylacetonate was investigated. Its preparation from zirconium(IV)alcoholate had been attempted earlier, but had failed. Dry benzene (110g) was poured over 24.24g zirconium(IV)acetylacetonate with 37.6 grams of phenol. This was refluxed 3 hours at 172°F, but no visible change occurred, and cooling to 32°F produced no precipitate. In another batch, benzene was eliminated to permit refluxing at a higher temperature. At a reflux temperature of 284°F to 311°F, the color changed successively from yellow to red-brown, to dark brown, then to tarry black--the color of the residue. The temperature in the second trial was too high, inasmuch as zirconium(IV)acetylacetonate is reported to begin decomposing at 257°F. Further investigation is in progress.

ZIRCONIUM(IV)PHENOXYALDEHYDE RESIN

The resin can be prepared by heating a mixture of zirconium(IV) phenoxide and paraformaldehyde. The reaction is rapid and difficult to stop in the "A" stage, therefore, the addition of a solvent is advisable. Zirconium(IV)phenoxide is readily dissolved in several solvents, such as dioxane, tetrahydrofuran, benzene, phenol, etc., and paraformaldehyde is soluble in water, ethyl alcohol, ether and possibly other solvents. The resin was successfully prepared by using dioxane or tetrahydrofuran as a solvent for the phenoxide. Paraformaldehyde is not soluble in these solvents; therefore, it was classified to pass through a 250 mesh screen and added as a

solid. During refluxing, the small, solid particles of paraformaldehyde are in constant motion. The phenoxide in solution appears to react with the surface of the aldehyde and by using a small particle size, the reaction time at the proper temperature is short enough that essentially all the aldehyde is consumed. The particles remaining are generally very small and have a slow settling rate in the resin. Resin prepared in this manner is controllable, and the reaction can be stopped at any desired viscosity or degree of polymerization. The resin can readily be thinned to impregnate glass cloth. One disadvantage is that its shelf life is only one to two weeks. The zirconium(IV)phenoxide is so reactive with paraformaldehyde that the mixed dry powders will polymerize at room temperature over a period of four weeks.

Since the shelf life of metal phenoxyaldehyde resins is so short, a titanium(IV)phenoxyaldehyde resin was prepared, substituting Methyl Formcel for paraformaldehyde. Methyl Formcel contains 46.5% aldehyde and 9% water, the remainder being methanol. A laminate was made, and further investigations are in progress. The shelf life is greater than four months.

A novolac prepolymer was prepared by adding half the stoichiometric amount of paraformaldehyde (22.5g) to zirconium(IV) phenoxide (11.5g) in solution with dioxane (150cc). This was refluxed five hours. A resin was obtained, but about half of it solidified on the sides of the flask. When stirred, the resulting prepolymer was lumpy, with solid particles suspended in the resin. Further effort will be expended to improve the resin.

LAMINATES

Laminates were prepared from Style 181 glass cloth with titanium (IV) and zirconium(IV)phenoxyaldehyde resins. A 1/8" X 8" X 10" laminate containing 30% resin was prepared with titanium(IV)-phenoxyaldehyde. The laminate was cured one hour at 300°F and 1000 psi pressure. A high resin flow was obtained but the laminate showed no starved areas. Physical properties (tensile and flexural strengths) of four specimens were determined with average values of 36,250 and 11,000 psi respectively.

The flexural specimens failed in transverse shear with a low value. In comparison, the ultimate tensile strength of zirconyl(IV) phenoxyaldehyde laminate (181 glass cloth) was 38,000 psi. Unfortunately the laminate size was inadequate to obtain flexural test specimens.

CONSTRUCTIVE PYROLYSIS

A laminate prepared from zirconium(IV)phenoxyaldehyde and Refrasil cloth (40% resin) was supplied to NARMCO Industries, a Division of Telecomputing Corporation, for evaluation by pyrolysis.

Refrasil (high silica glass) is leached glass cloth, and has very low tensile strength, thus the laminate also had low strength. Two more laminates were prepared, using titanium(IV)phenoxyaldehyde resin one with Style 181 glass cloth, the other with carbon cloth. Laminates had been investigated by MARMCO for constructive pyrolysis (in an inert atmosphere) to improve the oxidation resistance at elevated temperatures. The results of their studies indicated that "E" glass is superior to Refrasil for the resins in laminate production, since it is softer at 1800°F than quartz or Refrasil and is, therefore, more reactive chemically.

The flexural strength of the laminate of titanium(IV)phenoxyaldehyde resin with "E" glass was much higher after pyrolysis than that of the zirconium(IV)phenoxyaldehyde-Refrasil laminate, though they were of nearly the same strength before pyrolysis. The titanium in the resin forms a silicate with Refrasil in a rocket nozzle during test firing: hence, it would be expected to react with "E" glass during pyrolysis, resulting in a stronger laminate. The weight loss of the titaniumphenoxyaldehyde laminate also showed a pronounced increase in oxidation resistance at 650°F. Volume resistivity of both laminates was very low after the 1800°F pyrolysis. Resistivity of the carbon-cloth laminate, being initially low showed only insignificant change. In general, the over-all oxidation resistance of these laminates was not materially improved by pyrolysis at 1800°F.

ZIRCONIA FIBERS

Several nozzles were molded to shape from zirconium(IV)phenoxyaldehyde resin and zirconia "A" refractory fibers. Rocket motor test firings comparing this molding with one of zirconia fibers and 91LD resin indicated poor resistance to erosion for either nozzle. Since the 91LD resin used with several fiber reinforcements showed some merit in a nozzle environment it has been concluded that the crushing of the zirconia fibers during the molding process essentially yielded a powder filled molding known to be notoriously poor when used as a nozzle insert. A nozzle insert prepared with tantalum wire screen and zirconium hydride as a filler, however, actually reduced in throat size after test firing under the same conditions.

ZIRCONIUM-TANTALUM CARBIDE (IN SITU) LAMINATE

In the pyrolysis of the rocket nozzle inserts containing the polymers of various chemical elements, the solid residue has been the oxides as anticipated on the basis of enthalpy or free energy of formation values of the oxides from their elements. The carbides of the metals have lower negative free energy than the oxides. Because of the combined oxygen present in the inorganic-organic polymer and the apparent impossibility of preparing the polymer with a metal-to-carbon bond, the production of carbides through pyrolysis requires that a method be devised for removing the oxygen by an oxygen scavenger.

A laminate (1" X 1" dia.) was prepared by incorporating woven tantalum wire in zirconium(IV)phenoxyaldehyde resin with sufficient zirconium hydride to combine with all the oxygen in the the resin to form ZrO_2 and with enough excess to form a TaC-to-ZrC ratio of 4:1 (wt.) during destructive pyrolysis on the blast-exposed surfaces of a rocket nozzle. The hypothetical formation of the carbides in the 4:1 ratio is desired because, in that ratio, TaC and ZrC have one of the highest melting points of any known materials. Further it is believed that the metal screen will serve as an oxygen scavenger allowing the carbides to form on the exposed metal surface if an excess of carbon is present. Finally, the use of tantalum and zirconium fibers as reinforcing agents are of prime interest because of their extremely high melting temperatures and potential for supplying high melting compounds. The zirconium hydride in decomposition will produce an active form of elemental zirconium along with active hydrogen. Since the carbon monoxide and carbon dioxide would probably not form, because of their low enthalpy values, the hydrogen from the hydride would be released at the surface of the nozzle where it would absorb a considerable amount of heat as a result of its high heat capacity. Too high a rate of hydrogen evolution could lead to excessive intumescence of the nozzle material, whereas too low a rate would reduce the cooling effected by heat absorption by the hydrogen. In either event, the erosion rate would tend to increase. The optimum rate of hydrogen release would therefore, be an intermediate value, the net result of which would be the lowest erosion rate. Other metal hydrides may prove to be superior to zirconium hydride in that they may release hydrogen gas at more nearly the optimum rate.

BIBLIOGRAPHY

MARMCO Research and Development Division of Telecomputing Corp. Research and Development of High Temperature Structural Adhesives by S. H. Ruetman, W. V. Wrasidlo and H. H. Levine, MARMCO, November 30, 1962 (Contract NOw 61-0254-c) . UNCLASSIFIED

Prasad, Sarju and Upadhyaya, Railash. Nath, "Chemistry of Vanadium" JOURNAL OF THE INDIAN CHEMICAL SOCIETY, Vol. 37, No. 9, (Sept. 1960), p543 .

U. S. Naval Ordnance Test Station, Some Results of Applied Research On Materials and Processes for Providing Heat-Resistant Rocket Components, by Engineering Department, China Lake, Calif., NOTS, June 1961, (NAVORD Report 6390, NOTS TP 2666, Supplement 2) . CONFIDENTIAL

U. S. Naval Ordnance Test Station, Some Results of Applied Research On Materials and Processes for Providing Heat-Resistant Rocket Components, by Engineering Department, China Lake, Calif. NOTS, June 1962, (NAVORD Report 6390, NOTS TP 2863, Supplement 3) . CONFIDENTIAL

PHYSICAL PROPERTIES OF THE PALLADIUM-HYDROGEN SYSTEM

A. I. Schindler
U. S. Naval Research Laboratory
Washington, D. C.

I. Introduction

It is well known that palladium can absorb large amounts of both hydrogen and deuterium. At room temperature palladium can absorb more than 1000 times its own volume of hydrogen (or deuterium) which is equal to a hydrogen (or deuterium) to palladium atom ratio of unity. A two-phase region begins at low hydrogen concentrations (H/Pd≈0.03) and persists up to an atom ratio of approximately 0.6. The two phases consist of α-Pd, which has the same structure and approximately the same size as pure palladium, and β-Pd, which also has the cubic structure, but is 10% larger in volume. Beyond an atom ratio of 0.6 only the β-Pd phase is detected.

Worsham, Wilkinson and Shull(1), using neutron diffraction techniques, have shown that in the β-Pd phase the hydrogen atoms occupy octahedral sites of the face-centered cubic lattice. See Fig. 1. The β-Pd phase is quite stable, but the hydrogen atoms can be removed by heating or by oxidation. The completeness of the recovery of the α-Pd phase has been somewhat obscured because of the lack of agreement between various investigators. Some have reported complete recovery after desorption(2), while others(3) have reported the persistence of a small residual lattice expansion.

Well known room temperature properties of Pd-H include the linear decrease of the magnetic susceptibility and the almost linear increase of the electrical resistivity as a function of the hydrogen content. The fact that the magnetic susceptibility becomes negative and levels off at an atom ratio of 0.6, coupled with the standard rigid band

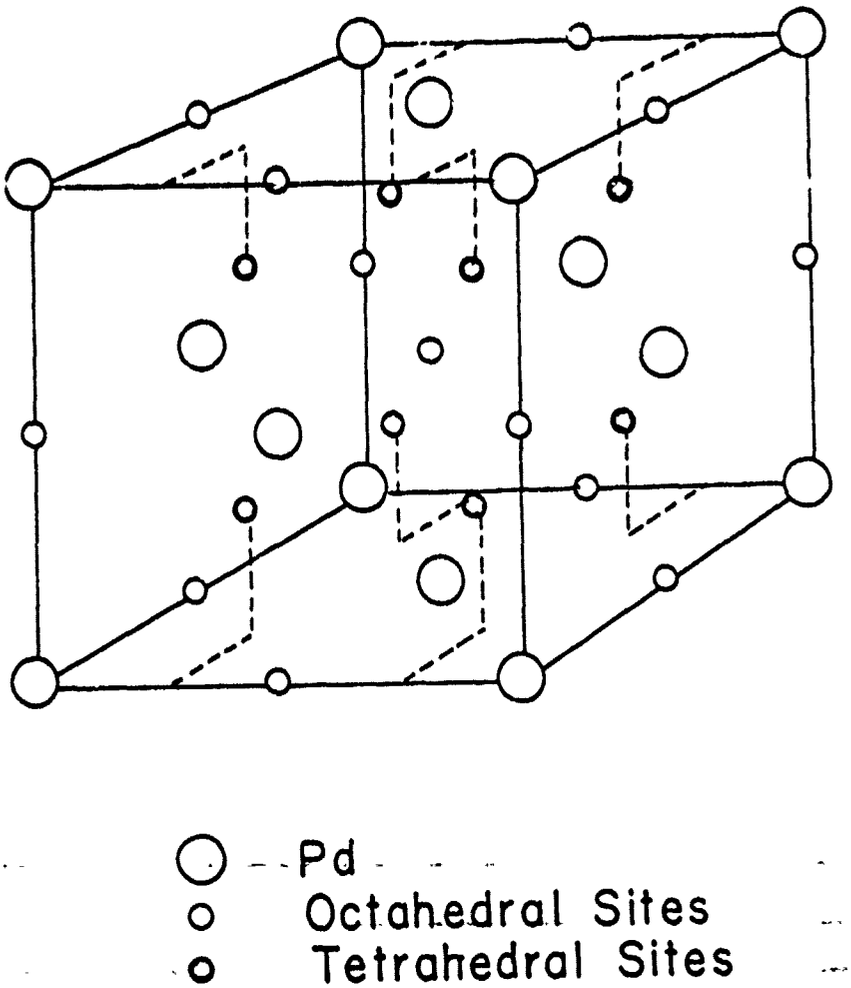


Fig. 1 - Model showing the possible positions of hydrogen and deuterium in the palladium lattice for ß-phase Pd-H and ß-phase Pd-D

assumption that palladium has 0.6 holes per atom in the d-band, led Mott(4) to suggest that when the hydrogen atoms were absorbed, they went into the lattice completely ionized with the stripped electrons filling the holes in the d-band. Although this model seems to be valid, it is based upon measurements made on a two-phase system. An alternate explanation for the linear decrease in magnetic susceptibility with hydrogen content can be given. The magnetic susceptibility of ß-Pd is found experimentally to be small and negative, and this does agree with Mott's hypothesis. Since the measurement of magnetic susceptibility yields a value which is an average over volume, the linear decrease in the susceptibility is more likely due to the proportion of α-Pd to ß-Pd decreasing linearly with

hydrogen content. Clearly some other experimental evidence to support Mott's model is in order.

With this aim in mind a series of experiments was begun which is still in progress. Measurements of the electrical resistivity, thermoelectric power, and lattice parameter of palladium were made as a function of temperature after charging with hydrogen(5), and again following desorption(6). In the latest series of experiments, the temperature dependence of the neutron scattering of palladium containing absorbed hydrogen and deuterium(7) is being examined.

II. Electrical Resistivity and Thermoelectric Power of Pd-H

a. Background

The original phase of this investigation consisted in making electrical resistivity and thermoelectric power measurements of Pd-H as a function of hydrogen concentration and temperature. In pure palladium the electrical resistivity versus temperature curve is concave downwards above 77°K. This is a normal consequence of the temperature dependence of the s-d scattering. Assuming a normal inverted d-band, the resistivity, ρ, can be written(8)

$$\rho(T) = AT(1+2\alpha\gamma T)\left[1-\frac{\pi^2}{6}\left\{\frac{kT}{E_1-\zeta}\right\}^2\right] \qquad (1)$$

where AT arises from the normal phonon scattering and is proportional to the density of states of the d-band, $N_d(\zeta)$, evaluated at the Fermi energy, ζ. The second part of the term, $(1+2\alpha\gamma T)$, arises from the lattice expansion; α is the coefficient of expansion, and γ is the Grüneisen constant. The remaining portion of the term, in the square brackets, is a consequence of the temperature dependence of the s-d scattering, where $(E_1-\zeta)$ is the width of the unfilled portion of the d-band. For a transition metal like palladium, the term in the square brackets, which is dominant, decreases with increasing temperature and results in the observed concave downward variation of the resistivity with temperature. If in ß-Pd, as postulated by Mott, the d-band of palladium is completely filled by the electrons stripped from the hydrogen atoms, one would obtain for the resistivity,

$$\rho(T) = AT(1+2\alpha\gamma T) \qquad (2)$$

Consequently, the temperature dependence of the resistivity of Pd-$H_{0.6}$ should either be linear or be slightly concave upwards (depending on the magnitude of $2\alpha\gamma$).

In spite of the inadequate theory of the thermoelectric power, it was also felt desirable to make such measurements as a function of temperature and hydrogen content. At high temperatures (T greater than the Debye temperature, θ_D) the simple formula for the thermoelectric power, S, is

$$S = \frac{\pi^2 k^2 T}{3|e|} \left.\frac{\partial \ln\rho}{\partial E}\right|_{E=\zeta} \qquad (3)$$

Again assuming the dominance of the s-d scattering on the resistivity and a normal inverted parabolic d-band, Eq. (3) reduces to

$$S = -\frac{\pi^2 k^2 T}{6|e|(E_1-\zeta)} \qquad (4)$$

Equation (4) would predict that S should become increasingly negative as the d-band holes are filled ($[E_1-\zeta]\rightarrow 0$).

b. Experimental Procedure

The electrical resistivity and thermoelectric power measurements were made using palladium wires approximately 0.015" in diameter. Each sample was electrolytically charged with hydrogen in a sulfuric acid solution using a platinum wire as the anode and the sample as the cathode. Hydrogen concentrations were determined by vacuum fusion procedures. Thermoelectric power measurements were made relative to a pure palladium wire (99.99% Pd), and the absolute thermoelectric power was determined from this data using the results of Borelius and coworkers(9). The details of the measurements have been reported elsewhere(5) and will not be repeated here.

c. Experimental Results

The electrical resistivity of palladium containing various amounts of hydrogen is plotted as a function of temperature in Fig. 2. Only the sample represented by the uppermost curve is single phase (β-Pd). Note that for this sample, there is a maximum in the resistivity at approximately 50°K, a minimum at 90°K, and the curve is concave upwards. As the hydrogen concentration is

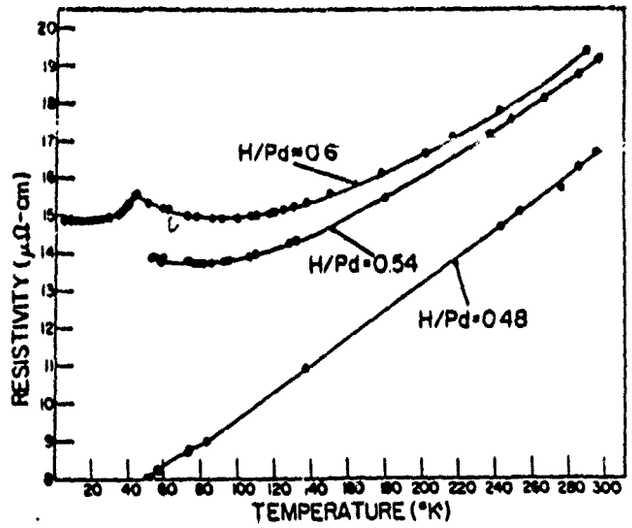


Fig. 2 - The variation of the electrical resistivity as a function of temperature for pure palladium containing various amounts of absorbed hydrogen

decreased, the maximum shifts to higher temperatures and the minimum to lower temperatures. For a hydrogen to palladium atom ratio of 0.48 (two-phase) an almost linear variation in the resistivity with temperature is observed. Although this latter behavior was expected, for this sample it is probably only a fortuitous result. It probably represents nothing more than the average temperature dependence of the two phases present, viz., concave up for α-Pd and concave down for β-Pd. In addition, it is found that the resistivity has been significantly increased by the absorption of hydrogen. At 4°K, the uppermost curve has a resistivity 75 times that of pure palladium at the same temperature. This would indicate that the increase in scattering arising from the hydrogen ions more than compensates for the decrease in scattering resulting from d-band filling. In fact, the contribution to the resistivity from the hydrogen ion scattering is almost two orders of magnitude larger than the s-d scattering contribution.

Figure 3 is a plot of the thermoelectric power relative to palladium as a function of temperature and hydrogen content. At H/Pd equal to 0.11 a very wide and shallow minimum is found in the vicinity of 65°K. As the hydrogen concentration is increased, this minimum becomes more pronounced and shifts toward lower temperatures. At H/Pd equal to 0.70, the minimum occurs at 50°K. It should

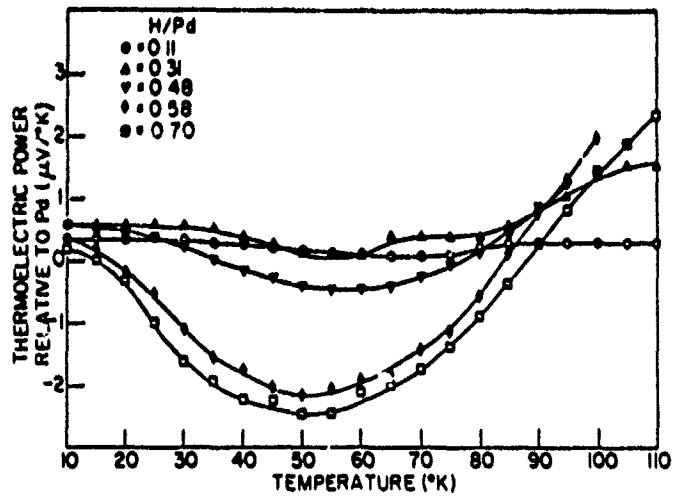


Fig. 3 - The variation of the thermoelectric power relative to pure palladium as a function of temperature for palladium containing various amounts of absorbed hydrogen

also be noted that the samples which are in the two-phase region (H/Pd=0.11, 0.31, 0.48) are grouped together with a shallow minimum while the single-phase samples (H/Pd=0.50, 0.70) form a second group with a pronounced minimum. From the similarity in behavior of the resistivity maximum and the thermoelectric power minimum, it would appear that both anomalies have a common origin.

The absolute thermoelectric power at various temperatures is plotted in Fig. 4 as a function of composition. The temperatures shown are in degrees Kelvin. These results have been obtained using the relative thermoelectric power data obtained here in conjunction with the absolute thermoelectric power data of pure palladium of Borelius, et al.(9). The data shown for the composition H/Pd=0 is, in fact, that of Borelius. It should be noted that the curves for temperatures between 70°K and 110°K have been displaced upwards for the sake of clarity. At 10°K a rather weak maximum is found for the absolute thermoelectric power between H/Pd=0.4 and 0.5. As the temperature is increased, the absolute thermoelectric power of samples containing the low concentrations of hydrogen increase more rapidly than those containing the high hydrogen concentrations. Consequently, between T=15°K and T=85°K, the data shows a rapid fall-off at H/Pd=0.6. At temperatures greater than 70°K, the absolute thermoelectric powers for samples having H/Pd greater than 0.48 continue to increase, while those of samples having

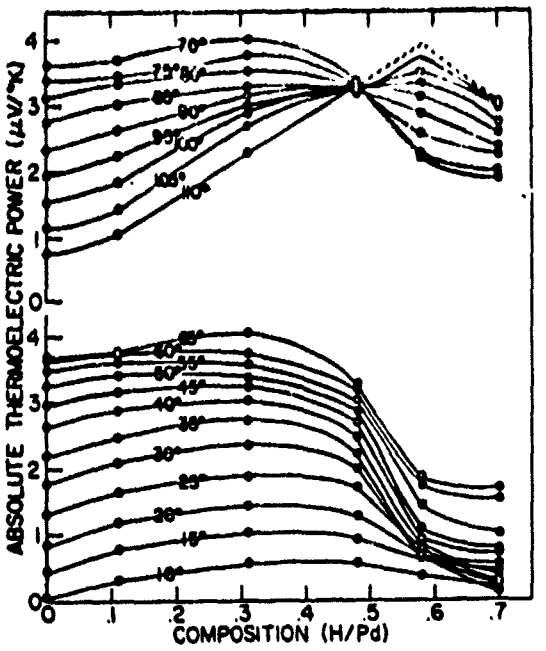


Fig. 4 - The variation of the absolute thermoelectric power as a function of temperature and hydrogen content for palladium containing absorbed hydrogen

H/Pd less than 0.48 begin to decrease. At temperatures greater than 95°K, a maximum in the absolute thermoelectric power is found at H/Pd=0.6.

c. Discussion

While the results of these measurements do not conflict with Mott's hypothesis of d-band filling via electrons from the ionized hydrogen, neither do they offer a clear verification. In partial verification we have:

(1) The appearance of the concave-upwards nature of the electrical resistivity with temperature for H/Pd=0.54 and 0.6. In one case, H/Pd=0.48, an almost linear variation was found. It should be pointed out, however, that this linear variation was found for a two-phase sample.

(2) The sharp fall-off with composition of the low temperature absolute thermoelectric power at H/Pd=0.6 (again relying upon the validity of the standard thermoelectric power theory).

In partial disagreement we have the maximum in the absolute thermoelectric power at H/Pd~0.6 for temperatures greater than 95°K.

In addition a new phenomenon has been found: a maximum in the electrical resistivity and a minimum in the thermoelectric power in the region of T~50°K for single-phase β-Pd. Several possible models can be suggested as the origin of these anomalies. These include a phase transformation, an atomic disordering, or an antiferromagnetic transition, occurring at 50°K. Since neutron diffraction can be used to detect all three types of transitions, such measurements were made.

III. Neutron Diffraction Studies

a. Background

Neutron diffraction examinations of palladium containing either absorbed hydrogen or deuterium have already been made at room temperature by Worsham, et al.(1), and more recently by Bergsma and Goedkoop(10). Their data showed, at room temperature, that no new diffraction lines appeared, which indicated that the absorbed atoms scattered neutrons coherently with palladium. Worsham, et al., found that the results could be analyzed on the basis of a NaCl structure with either the hydrogen or deuterium atoms occupying octahedral sites in the face-centered cubic structure of the palladium lattice. For such a structure, $F_{hk\ell}$, the structure factor for an $(hk\ell)$ reflectron can be written,

$$F_{hk\ell} = f_{Pd}\exp(-b_{Pd}\sin^2\theta/\lambda^2) \pm (H/Pd) f_H \exp(-b_H \sin^2\theta/\lambda^2) \quad (5)$$

+ for h, k, ℓ all even, - for h, k, ℓ all odd

where the f's are neutron scattering amplitudes, b is the Debye-Waller temperature factor, 2θ is the scattering angle, and λ is the neutron wavelength. A similar equation can be written for palladium containing absorbed deuterium. Since f_{Pd} has the same sign as f_D, the reflections with even indices will be enhanced and the reflections with the odd indices will be attenuated for palladium containing absorbed deuterium. On the other hand, f_{Pd} has a sign which is opposite to that of f_H, so for palladium containing absorbed hydrogen the reflections with the odd indices will be enhanced and the reflections with the even indices will be attenuated. From their data, Worsham, et al., concluded that the absorbed hydrogen or deuterium atoms occupied only

octahedral sites. In addition, they were able to extract the Debye-Waller temperature factor (b), from which they were able to calculate the root mean square vibrational amplitude, $\sqrt{(\overline{u^2})}$, of the various species of atoms.

$$b = 8\pi^2\sqrt{(\overline{u^2})} \qquad (6)$$

Their data yielded:

Element	Sample	$\sqrt{(\overline{u^2})}$
Pd	Pd	0.069 Å
Pd	$Pd\text{-}H_{0.7}$	0.071
Pd	$Pd\text{-}D_{0.66}$	0.064
H	$Pd\text{-}H_{0.7}$	0.23
D	$Pd\text{-}D_{0.66}$	0.20

Bergsma and Goedkoop obtained data which agreed quite well with these results:

Pd	$Pd\text{-}H_{0.6}$	0.10±0.01 Å
H	$Pd\text{-}H_{0.6}$	0.24±0.02

In addition Bergsma and Goedkoop found the Debye temperature, θ_D, to be approximately 300°K for $Pd\text{-}H_{0.6}$ (which is not much larger than $\theta_D \sim 275$°K for pure palladium).

In order to determine the origin of the anomalies in the electrical resistivity and thermoelectric power, similar neutron diffraction experiments were performed as a function of temperature from 4.2°K to room temperature. For normal metals, the integrated intensities of all the diffraction lines should increase as the temperature is lowered. This can easily be shown.

$$I_{hk\ell} = \frac{kjF^2_{hk\ell}}{\mathrm{Sin}\theta\mathrm{Sin}2\theta} \qquad (7)$$

where I is the integrated intensity under the diffraction peak, j is the multiplicity of the diffraction line, and k is an equipment constant. From Eq.(5), Eq.(6) and Eq.(7), one sees that as the temperature is lowered, $\sqrt{(\overline{u^2})}$ should decrease, F should increase, and consequently I should increase. If a phase transition or an antiferromagnetic transition occurred at the critical temperature,

new diffraction lines would be expected to appear superposed on the normal temperature dependence.

b. Experimental Procedure

For the neutron diffraction experiments, sintered right circular cylinders of pure palladium sponge, about 3/4" diameter by 1" high, were employed. These cylinders were placed in quartz capsules and were repeatedly heated and cooled in the presence of purified tank hydrogen or deuterium. After the final cooling to room temperature, the capsules were sealed off. Each sample was placed, in turn, in a helium dewar arranged for neutron diffraction experiments. By employing baths of liquid nitrogen and liquid helium, and by pumping on the nitrogen, measurements could be made at 4.2°K, 50°K and 77°K. A room temperature, 300°K, measurement was also made. The neutron wavelength used was 1.09 Å, and was obtained by Bragg reflection from a lead crystal. In the early stages of this experiment, the reactor power was 100 kilowatts and the beam intensity was 10^6 neutrons/minute. This was later increased by a factor of ten as a result of an increase in reactor power. The initial observations consisted in scanning the diffraction pattern between counter angles of 5° to 80°. In the later measurements the temperature dependence of the four major diffraction lines, (111), (113), (200), and (220), were carefully observed.

c. Experimental Results

During the initial measurements at low reactor power level, 100 KW, it was observed that no new diffraction lines occurred between 4.2°K and 300°K. However, it was observed that the temperature dependence of the (111) line for Pd-H was clearly anomalous. As the temperature was lowered, the intensity was found to increase initially and then to decrease at 50°K. This decrease in intensity appears to be related to the observed electrical resistivity maximum and thermoelectric power minimum found earlier.

When the power level of the reactor was increased to 1 megawatt, it became possible to clearly resolve the (111), (113), (200), and (220) lines. The integrated intensities of these lines at various temperatures are shown in the following table. For the deuterium-palladium sample it was not possible to separate the (113) contribution from the (222) contribution, hence no value for $I_{(113)}$ is given.

Observed Intensities for Pd-H

Reflection	300°K	77°K	50°K	4.2°K
(111)	2811 ± 10	3585 ± 40	3064 ± 30	3599 ± 40
(200)	461 ± 5	396 ± 3	352 ± 6	687 ± 7
(220)	510 ± 5	552 ± 5	550 ± 5	657 ± 6
(113)	1855	2810	1826	2789

Observed Intensities for Pd-D

Reflection	300°K	77°K	50°K	4.2°K
(111)	412 ± 4	402 ± 4	626 ± 6	480 ± 5
(200)	2810 ± 20	2354 ± 22	2346 ± 22	2352 ± 22
(220)	1820 ± 20	1912 ± 20	1970 ± 20	1959 ± 20

From these results the root mean square vibrational amplitude can be obtained. At room temperature, we find that for palladium $\sqrt{(\bar{u}^2)} = 0.080$ Å, and for hydrogen $\sqrt{(\bar{u}^2)} = 0.239$ Å. This is in substantial agreement with the earlier quoted results. It is now possible to clearly observe the anomalous temperature dependences at 50°K of the various reflections. This can best be done by comparing the intensities for two temperatures which are relatively close together, i.e., I at 77° and at 50°K. For normal metals, one would not expect large differences in the vibrational amplitudes, and consequently in the observed intensities, to occur between these two rather low and substantially equal temperatures. At most one would observe a very small increase in intensity from 77°K to 50°K. The observed variation in going from 77°K to 50°K is tabulated as follows:

	$I_{(111)}$	$I_{(113)}$	$I_{(200)}$	$I_{(220)}$
Pd-H	decrease	decrease	increase	constant
Pd-D	increase		constant	increase

As we continue to go down in temperature, a more normal and positive change is found in all the observed reflections for the Pd-H sample, while a mixed behavior is found for the Pd-D samples.

d. Discussion

It is clear from the data that the normal temperature dependence is not found for the integrated intensities. Nor can the data be explained by the occurrence of a phase transformation or any of the other possibilities suggested earlier. However, it can be shown that if some fraction of the absorbed atoms occupy interstitial tetrahedral sites, the observed behavior can be explained. These sites are located along the body diagonals at the positions (1/4, 1/4, 1/4), (3/4, 3/4, 3/4), plus the face-centered operations. See Fig. 1. The diffraction equations now become:

$$F_{(hk\ell)} = f_{Pd}\exp(-b_{Pd}\mathrm{Sin}^2\theta/\lambda^2) - n(H/Pd) f_H \exp(-b_H \mathrm{Sin}^2\theta/\lambda^2) \quad (8)$$

for h,k,ℓ all odd

and

$$F_{(hk\ell)} = f_{Pd}\exp(-b_{Pd}\mathrm{Sin}^2\theta/\lambda^2) \pm (1-n)(H/Pd) f_H \exp(-b_H \mathrm{Sin}^2\theta/\lambda^2)$$
$$+ n(H/Pd) f_H \exp(-b_H \mathrm{Sin}^2\theta/\lambda^2) \quad (9)$$

for h,k,ℓ all even

The plus sign in Eq.(9) is taken for $h+k+\ell=4N$, while the sign is taken for $h+k+\ell\neq 4N$, with N an integer. n is the fraction of hydrogen atoms occupying octahedral sites. A similar set of equations can be written for palladium containing absorbed deuterium. For the two observed reflections with even indices Eq.(9) becomes:

$$F_{(220)} = f_{Pd}\exp(-b_{Pd}\mathrm{Sin}^2\theta/\lambda^2) + (H/Pd) f_H \exp(-b_H \mathrm{Sin}^2\theta/\lambda^2) \quad (10)$$

and

$$F_{(200)} = f_{Pd}\exp(-b_{Pd}\mathrm{Sin}^2\theta/\lambda^2)$$
$$+ (2n-1)(H/Pd) f_H \exp(-b_H \mathrm{Sin}^2\theta/\lambda^2) \quad (11)$$

It can thus be seen that if some fraction of the hydrogen atoms, (1-n), shift from octahedral to tetrahedral sites, $F_{(220)}$ will be unaffected, $F_{(111)}$ and $F_{(113)}$ should decrease, and $F_{(200)}$ will increase. Such atom movement would thus result in a decrease of $I_{(111)}$ and $I_{(113)}$, an increase in $I_{(200)}$ and would leave $I_{(220)}$ unaffected. This is

precisely what has been observed. Since the sign of f_D is opposite to that of f_H, a similar but opposite behavior would be expected. One should expect an increase in $I_{(111)}$ and $I_{(113)}$, a decrease in $I_{(200)}$ and, again, no change in $I_{(220)}$ at the critical temperature. Here, the agreement of the model with the observation is not as good as in the Pd-H case. Experimentally we find the expected behavior only for the temperature dependence of the (111) diffraction line. This discrepancy might result from the fact that since the mass of the deuterium ion is greater than that of the hydrogen ion, the diffusion rate of deuterium in palladium is lower than for hydrogen in palladium. It would appear that some of the observed discrepancies found represent the fact that equilibrium has not been established during the measurement.

III Electrical Resistivity and Thermoelectric Power after Hydrogen Desorption

a. Background

In order to accommodate the hydrogen atoms on interstitial octahedral sites at room temperature, the palladium lattice expands linearly by approximately 3%. This expansion can be considered equivalent to an internal strain. As a result, desorption processes which are carried out at low temperatures may leave the palladium lattice in a highly strained condition. Under such circumstances one might expect large increases in the electrical resistivity. It has also been shown that the thermoelectric power is extremely sensitive to lattice imperfections(11). For these reasons, measurements of the electrical resistivity and thermoelectric power were made on samples of desorbed palladium. Finally to check on the reported hysteresis effect in the lattice spacing, precision lattice parameter measurements were made on annealed palladium and again on palladium after desorption.

b. Experimental Procedure

All measurements were made on 15 mil diameter, 99.995% pure palladium wires in the following conditions: unannealed, annealed, charged with hydrogen and following hydrogen desorption. The annealed condition was obtained by heating the samples in vacuum at 800°C. The samples were then charged with hydrogen by the electrolytic method described earlier. Desorption was accomplished by heating the hydrogen-charged samples in vacuum for periods of 1/2 hour or less, up to a maximum temperature of 800°C. The complete removal of the hydrogen was checked by vacuum fusion procedures and by X-ray examination.

c. Exporimental Rosults

Back roflection X-ray examinations woro mado on samplos previously charged with hydrogen and thon desorbod. Tho rosults obtained conclusivoly domonstrato the prosonco of a hysteresis effect in the lattice. An increase in lattice dimensions was found which is approximately 1 part in 4000, in agreement with earlier measurements. Similar measurements made with desorbed palladium, previously charged with deuterium, indicated a residual lattice expansion of 2-3 parts in 4000.

The thermoelectric power measurements are summarized in Fig. 5. The upper curve shows the variation of

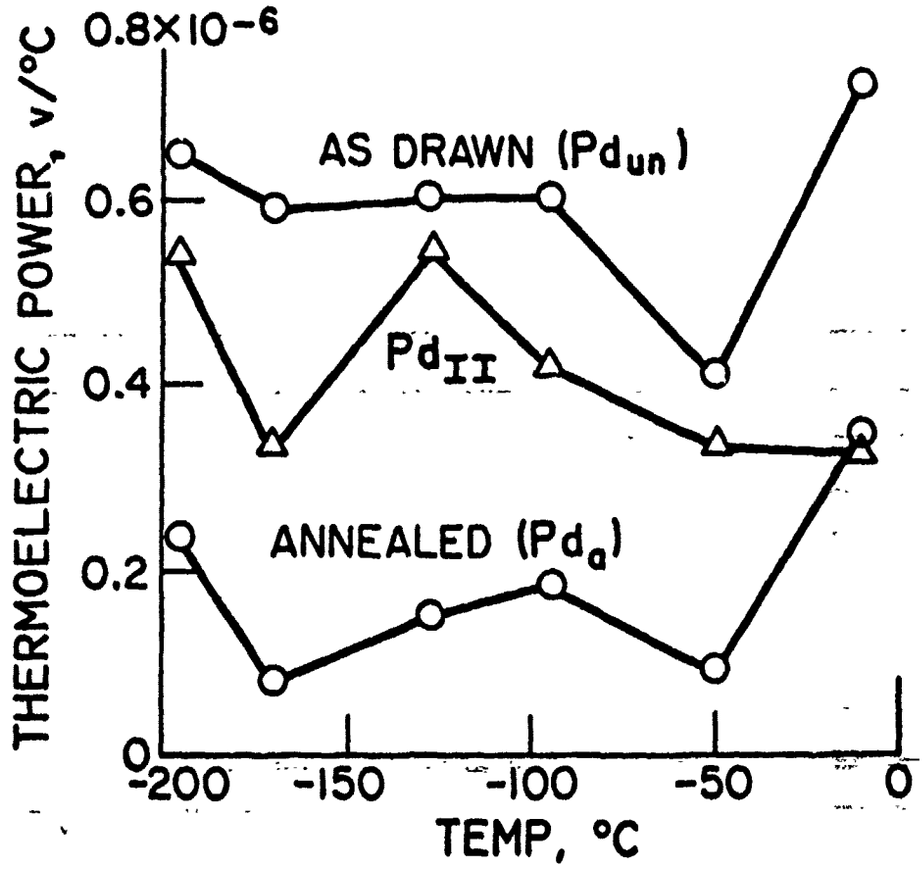


Fig. 5 - Shift of the thermoelectric power of desorbed palladium (Pd_{II}) from the annealed toward the unannealed value

the thermoelectric power with temperature for a sample of heavily cold-worked palladium (Pd_{un}), while the lowest curve is that for the same sample following a 900°C, 2-hour anneal. Between these curves is a plot of the thermoelectric power of a sample of palladium (after the sequence of hydrogenation, followed by a 300°C vacuum anneal to remove the hydrogen). An obvious shift from the annealed toward the cold-worked values is apparent.

Figure 6 shows the variation of resistance with temperature for Pd_{an}, annealed palladium, Pd_{II}, following desorption of hydrogen, and ß-phase palladium. The effects of absorbed hydrogen on the resistivity of palladium is given for completeness and is shown by the elevated points. It should be especially noted that contrary to expectation, Pd_{II}, at room temperature, has a resistance value 8% lower than annealed palladium. Denoting the difference in resistance of annealed palladium and Pd_{II} by ΔR, it is seen that ΔR decreases with decreasing temperature. In Fig. 7, which is a magnified view of the lower section of Fig. 6, $|\Delta R|$ is seen to go to zero and then to increase with decreasing temperature. At 4°K, Pd_{II} has a resistance value 5% higher than that of the same annealed sample. Values of the slope dR/dT were also obtained and are shown in Fig. 8. It is seen that over the entire range indicated, dR/dT is larger for the annealed Pd than for Pd_{II}.

d. Discussion

The electrical resistivity of a metal can be written:

$$\rho = \rho_o + \rho_T \qquad (12)$$

where ρ_o is the residual resistivity and ρ_T is the phonon contribution to the resistivity. The shift in the thermoelectric power of Pd_{II} toward the value for the cold-worked palladium would suggest that the process of absorption and desorption of hydrogen results in a lattice with residual imperfections. These imperfections in the lattice increase ρ_o, and this is observed, since Pd_{II} has a higher resistance than annealed palladium at 4.2°K. However, if the only difference in annealed palladium and Pd_{II} were the imperfections, the phonon contribution of the two states would be the same and the resistance of the desorbed sample should always be higher than the annealed sample. This, however, is not the case. If one writes $\rho_T = T(d\rho/dT)$, then (since $d\rho/dT$ for Pd_{II} is smaller than $d\rho/dT$ for annealed Pd), ρ_T for Pd_{II} is seen to be

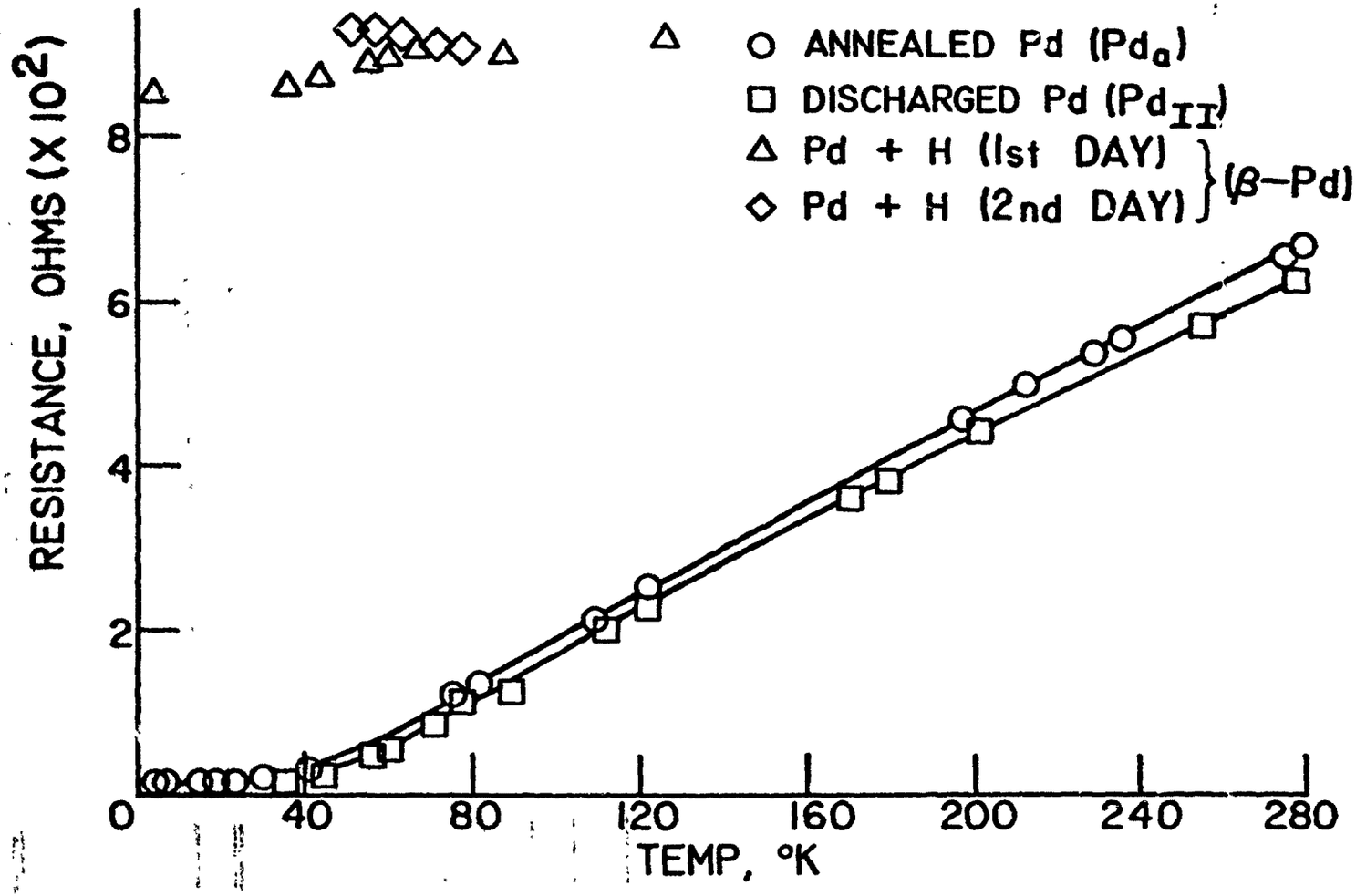


Fig. 6 - Effect of temperature on the resistance of palladium in the following states: annealed, charged with hydrogen, and after hydrogen desorption

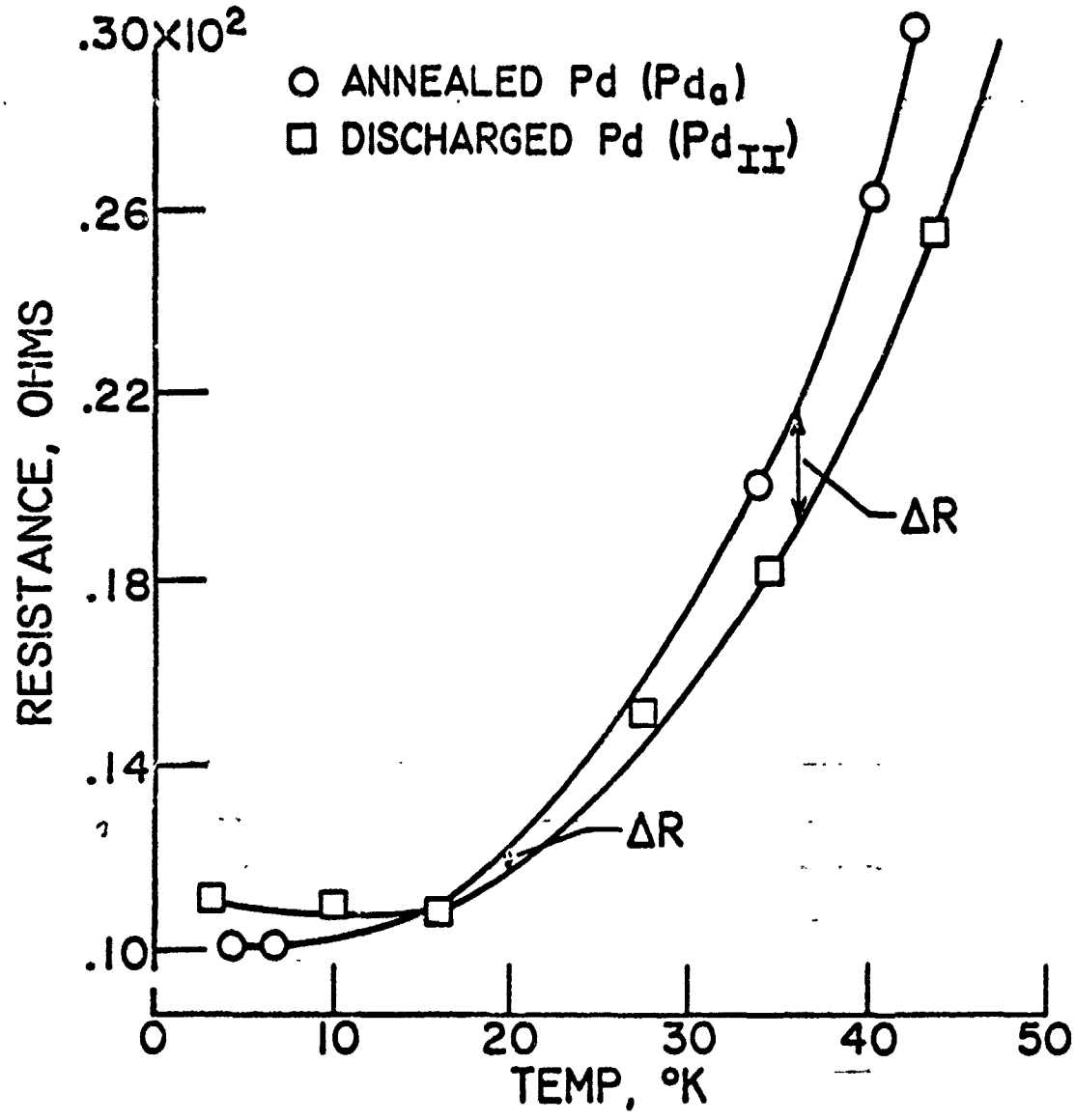


Fig. 7 - Enlarged view of the resistance versus temperature curves of annealed palladium and Pd_{II} showing the variation of ΔR with temperature

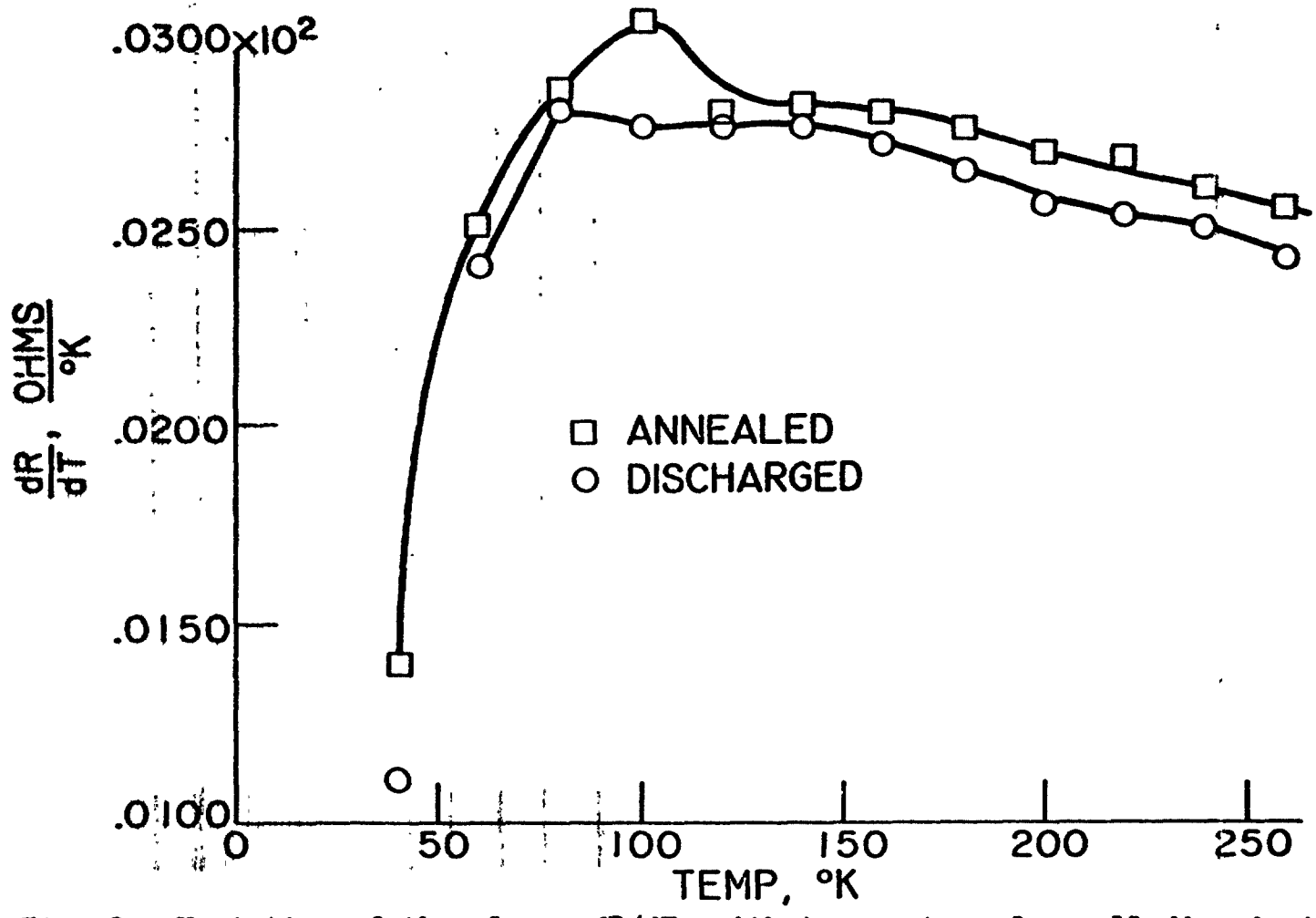


Fig. 8 - Variation of the slope, dR/dT, with temperature for palladium in the annealed state and after hydrogen desorption

lower than that of annealed palladium. As a result, the increase in ρ_o for Pd_{II} is more than compensated for by the decrease in ρ_T, and at room temperature Pd_{II} has a lower resistivity than that of annealed palladium.

Phenomenologically, the decrease in the phonon contribution to the resistivity can be described in terms of the modified lattice dynamics of the palladium after hydrogen desorption. The expanded lattice results in a stiffer lattice - the hardness has been observed to increase by a factor of 2 - and this would suggest an increase in the value of θ_D. In the case of a simple metal one can describe the resistivity in terms of θ_D.

$$\rho_T = \frac{T}{M\theta_D^2} \tag{13}$$

or

$$\Delta\left\{\frac{\partial\rho_T}{\partial T}\right\} = -\left\{\frac{2}{M\theta_D^3}\right\} \tag{14}$$

where M is the atomic mass. From the temperature dependence of the resistivity one can obtain the value of θ_D. The resistivity results shown above indicate an increase in θ_D of approximately 10°K. The observed increase in θ_D results in a decrease in $\partial\rho_T/\partial T$, as has been experimentally observed. However, it would be expected that expanding the lattice of a normal metal should result in a "looser" lattice and θ_D ought to decrease.

It is suggested that the expanded palladium lattice can cause a redistribution of s and d electrons in such a manner as to fill the d-band to a somewhat higher level. This would result in a small decrease in the density of states as well as in the width in energy of the unfilled portion of the d-band, $(E_1-\zeta)$. From Eq.(1) it can be seen that such a redistribution would result in a decrease in ρ_T because of a decrease in A as well as in the latter part of the term in the square brackets. If this model is correct, such physical properties as Hall coefficient, electronic specific heat and magnetic susceptibility should also be affected. Measurements of these properties are being planned.

ACKNOWLEDGMENTS

The results presented here represent the combined efforts of many people. The transport property investigations were carried out in cooperation with Mr. R. J. Smith

and Dr. E. W. Kammer, while the neutron diffraction experiments were made by Mr. G. A. Ferguson. The author is grateful to Mr. C. L. Vold for the X-ray measurements quoted, as well as for many helpful discussions concerning diffraction theory. He also wishes to thank Mr. G. Picklo and Mr. D. Walter for making the gas analyses of all the samples used in these experiments.

REFERENCES

1. J.E. Worsham, J.K. Wilkinson, and C.G. Shull, J. Phys. Chem. Solids, 3, 303 (1957).

2. F.A. Lewis, Metals Rev. 4, 132 (1960).

3. G. Chaudron, A. Portevin, and L. Moreau, Compt. Rend., 207, 235 (1938).

4. N.F. Mott, Proc. Phys. Soc., 47, 571 (1935).

5. A.I. Schindler, R.J. Smith, and E.W. Kammer, Proceedings of the Tenth International Congress of Refrigeration, Copenhagen, 1959 (Pergamon Press, New York, 1960), Vol. 1, p. 74.

6. R.J. Smith, A.I. Schindler, and E.W. Kammer, Phys. Rev. 127, 179 (1962).

7. G.A. Ferguson and A.I. Schindler, Rept. of NRL Progress, April, p. 42 (1962).

8. H. Jones, in Handbuch der Physik, edited by S. Flügge (Springer-Verlag, Berlin, 1956), Vol. 19, p. 267.

9. G. Borelius, W.H. Keesom, C.H. Johansson, and J.O. Linde, Amsterdam Roy. Acad. Sci. Proc., 33, 17 (1930).

10. J. Bergsma and J.A. Goedkoop, Physica, 26, 744 (1960).

APPLICATIONS OF RADIOCHEMISTRY TO SOLUTION OF RADIOLOGICAL PROBLEMS

N. E. Ballou
U. S. Naval Radiological Defense Laboratory
San Francisco, California

Radiochemical techniques have been extensively applied to the solution of radiological problems in the areas of radiological effects (including both early and late time dose rates and the influence of base-surge, fallout, detonation conditions, and fractionation on these) contamination-decontamination, and yields of nuclear weapon detonations. It is the purpose of this paper to describe the application of radiochemical techniques to determinations of those properties of products of nuclear weapon detonations which are of primary importance to radiological effects and contamination-decontamination and to determinations of weapon yields.

The earliest estimates of dose rate decay were obtained from calculations based on both theory and experimental data. (1) From this came the widely used approximation that the dose rate for the fission products from nuclear weapons detonations decayed proportionately to $t^{-1.2}$. More precise calculations showed that the exponent takes various values between about -1.0 to -1.6 for given time intervals in the range of one minute to three years after fission. (2) These calculations utilized radiochemically established half-lives and abundances (fission yields) of the fission product radionuclides; the activity of each radionuclide was computed for selected times after fission and the total activity at each of the time points was summed to give the gross fission product activity as a function of time. The contribution of each fission product radionuclide and element to the total fission product activity was also derived from these data.

In addition, the contributions of radionuclides produced by reactions of neutrons liberated in the detonation with bomb and environmental materials have been found to be significant in both fission and thermonuclear weapon detonations. These contributions have been quantitated by radiochemical means. Especially large deviations from decay rates of fission product mixtures have been

observed in material produced in thermonuclear detonations. Radiochemical analyses have shown that these differences can be ascribed largely to the neutron-induced radionuclides uranium-239 and neptunium-239 with smaller contributions from uranium-240, neptunium-240 and uranium-237. These radionuclides are produced by neutron reactions in uranium-238 and, in some cases, beta-decay of the product radionuclides. In thermonuclear detonations the contributions to the total activity by uranium-239 and neptunium-239 may be as high as about sixty percent at particular times after detonation while contributions from uranium-240 and neptunium-240 may go up to about twenty-five percent; uranium-237 contributions are less important, being no more than about ten percent. In fission weapon detonations only uranium-239 and neptunium-239 make significant contributions to the total activity. Their contributions may go up to about ten and twenty-five percent, respectively. It should be emphasized that large variations in these contributions can be expected because of the variety in design of nuclear weapons. Representative decay curves for products of thermonuclear and fission weapon detonations and for products of thermal neutron fission of uranium-235 are given in Figure 1 and are labelled, respectively, T, F and G. In nuclear detonations in seawater the predominant environmental neutron-induced radionuclide is sodium-24. Although the fraction of the disintegrations due to sodium-24 is no more than about 1% for fission weapon detonations, (3) its contribution to the dose rate is appreciably higher (about 4%) due to the very energetic gamma-rays associated with its decay. The decay curves of fission products and induced radionuclides from a deep underwater nuclear detonation are given in Figure 2.

Computations of gross fission product decay characteristics at times less than about 1 hour after fission are appreciably less reliable than those for later times. This is due to the incomplete data on fission product radionuclides present at these shorter times. Because of the operational importance of knowledge of such short time decay characteristics, two approaches have been used to obtain the needed data. One is based on theoretical predictions of fission product half-lives and abundances, (4,5) and the other is based on specially designed fast radiochemical measurements. (6,7,8,9) In the first case, disintegration rates for the products of thermal neutron fission of uranium-235 have been obtained for the time interval of one second to two hundred years after fission. In the second case, gamma-ray ionization rates and energy release rates of the products of thermal neutron fission of uranium-235 and of plutonium-239 have been obtained over the time interval of one second to about one day following fission.

In order to assist in the interpretation and prediction of radiological effects radiochemical measurements have been made on materials resulting from nuclear detonations occurring at several depths in shallow and deep water and on lagoon bottoms. (3,10,11,12, 13,14,15) Such measurements help to reveal the influence of these

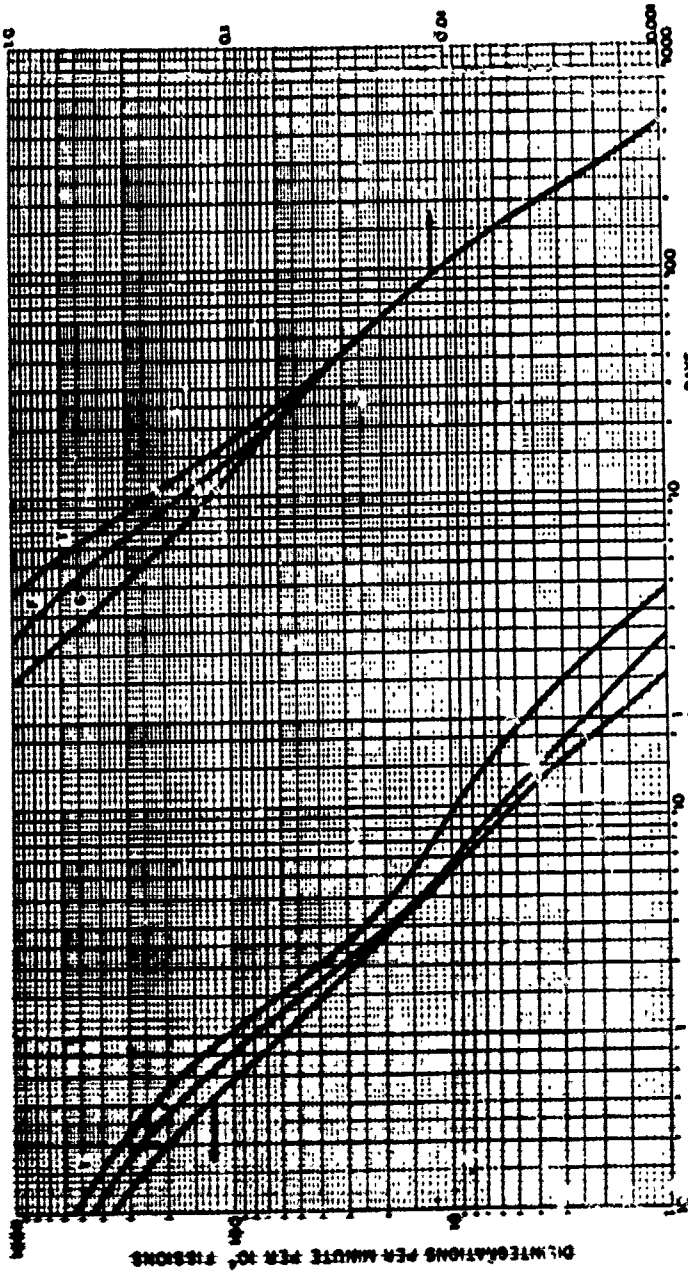


Figure 1 Representative Decay Curves of Products of Thermonuclear (T) and Fission (F) Weapon Detonations and of Thermal Neutron Fission of Uranium-235 (G)

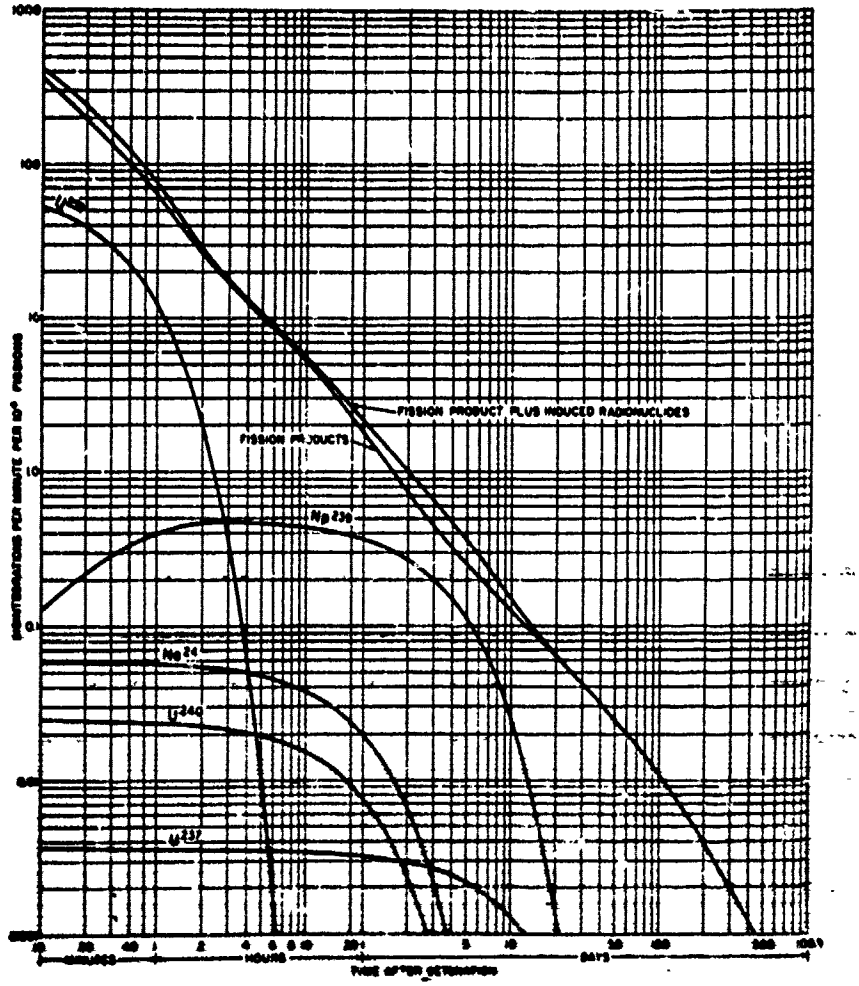


Figure 2 Representative Decay Curves of Products of Fission Weapon Detonations in Seawater

varying detonation conditions on dose rates and on the nature and extent of base-surge, fallout and fractionation. However, because of the quite limited number of appropriate nuclear detonations occurring and because of the very great expense and difficulty in performing adequate experimental measurements under immediate post-detonation conditions, it is essentially impossible to make satisfactory measurements on composition and characteristics of debris at short times after detonation. Consequently, laboratory measurements are required to establish the radionuclide composition and characteristics of the fission product mixture at early times after fission. Such laboratory experiments have been initiated and are successfully supplying necessary information.

Certain phenomena in sub-surface nuclear detonations have been observed to have major radiological consequences, e.g., generation of a base surge, entrapment of radioactive debris, fractionation. The base surge arising from subsurface nuclear detonations in water is of prime operational importance because of the high levels of radioactivity associated with it. Radiochemical analyses and decay measurements have shown the base-surge to be markedly different in radiochemical composition and radiation characteristics from the fission products of the detonation taken as a whole. Such non-representative distribution of constituents of bomb debris is designated "fractionation". This phenomenon has been found to be extensive and important in surface and subsurface detonations in both soil (16) and water. (10,11,12,13,17) Fractionation can be explained by the different chemical and physical properties of the radioactive elements constituting the nuclear debris. The simplest illustration of this comes from the behavior of rare gas fission products. For example, in a subsurface detonation in water the rare gas radionuclides initially formed will break the surface along with other gases and some of the non-volatile fission product elements. The latter will have been at least partially scrubbed out of the ascending bubble and will be further removed as the ejected water falls back into the ocean. However, the rare gas fission product elements, xenon and krypton, have a greater tendency to remain in the gas phase. Thus the gas phase becomes enriched in the rare gas radionuclides and their descendents. As a consequence of this the radiation characteristics and the composition of this portion of the nuclear debris is markedly altered.

The contamination behavior and the effectiveness of given decontamination procedures are affected by the chemical composition and the chemical state of the nuclear debris. (18,19) The association of the gross fission product mixture and of selected radionuclides with soluble, colloidal and particulate phases has been estimated theoretically (20), and it has been measured on material from laboratory simulation of underwater nuclear detonations (21) and on material collected from subsurface detonations in deep water and on lagoon bottoms. (3,11,12,22) General agreement has been observed among theoretical estimates and results from the simulation

experiments and from actual detonations. In a deep water detonation from 35-65% of the fission product activity was found to be in the soluble state, with the higher solubilities occurring at later times after detonation, i.e., 2 weeks or more. Distributions of total activity with particle size were determined in a deep water and in a lagoon bottom (coral) detonation. In the former there was a gradual decrease in fraction of activity associated with particles in the size range 0.002 μ to 20 μ. In the lagoon bottom event, activity remaining suspended at two hours after detonation was found to be associated with two groups of particles having sizes in the neighborhood of 0.002 μ and at about 10 μ.

Oxidation states have been found to be important in the contamination and decontamination behavior of certain elements.(18, 19) Examples are neptunium and iodine. A limited number of measurements of this property have been made on material from nuclear detonations. (3,11,23) Neptunium was found to be predominantly in the V and/or VI state and uranium was mainly in the VI state. Iodine was found to varying degrees in its different oxidation states depending on the age of the material.

The primary means by which the yield of a fission weapon is determined is by the use of radiochemical measurements. Other types of measurements are calibrated in respect to the radiochemically determined yield. In the radiochemical methods two principle quantities must be determined, i.e., (a) the number of fission events which produce the fission products in a collected sample and (b) the fraction of the bomb debris represented by the sample. From this is calculated the total number of fission events in the detonation and thus the yield.

The number of fission events represented by a sample is determined by measurements of selected fission product radionuclides whose fission yields are well known. Molybdenum-99 and neodymium-147 serve well for this purpose. The fraction of the bomb is determined from measurement in the sample of the amount of an element which was originally present in a known amount in the nuclear weapon. Uranium-235 and plutonium-239 are utilized for this purpose. Because of the presence of uranium in seawater and the great dilution of the weapon debris in a detonation, uranium-235 has not been found to be very useable in underwater detonations and primary use is made of plutonium-239.

In addition to the measurements of fission monitors and bomb fraction indicators a wide variety of other radiochemical analyses must be performed because of (a) the many nuclear reactions other than fission which occur in a nuclear weapon detonation, (b) the presence of more than one fissionable element in a weapon, and (c) fractionation. As examples of the perturbing affects of other nuclear reactions, one can consider the loss of plutonium-239 and uranium-235 by (n,γ) and (n,2n) reactions, the production of

plutonium-239 by neutron capture in uranium-238 and the production of plutonium-240 by double (n,γ) processes on uranium-238. Since there are some variations in fission yields of the fission monitors in fission of different nuclides, the fraction of fissions contributed by each of the fissile nuclides in a device must be established. One way to do this is to radiochemically determine relative amounts of radionuclides whose production is sensitive to the fissioning nuclide. Examples of such are silver-111, cadmium-115 and europium-156. Fractionation among the fission products or between the fission monitors and the bomb fraction indicators will clearly distort any weapon yield value. Analyses for such elements as barium-140, strontium-89, and zirconium-95 serve to establish the existence and extent of fractionation. If such occurs, corrections for it may be derived from empirically established fractionation correlations provided fractionation is not too severe. Mass spectrometric analyses are made also in order to establish isotopic ratios in uranium and plutonium. The resulting data provide information on the extent of various nuclear reactions occurring in those elements, in addition to permitting further interpretation of certain radiation measurements (e.g., alpha pulse height analyses of plutonium-239, 240).

In summary, the application of radiochemistry to the definition of properties of nuclear weapons debris has been described. Properties selected as of importance to radiological effects include dose rates, radionuclide composition (fission product and neutron induced radionuclides), fractionation, and physical and chemical states (solubilities and oxidation states). In addition, the radiochemical method for determination of fission weapon yields has been outlined.

References

1. K. Way, and E. P. Wigner, "The Rate of Decay of Fission Products", Phys. Rev. 73, 1318-1330 (1948).

2. H. F. Hunter, and N. E. Ballou, "Simultaneous Slow Neutron Fission of U^{235} Atoms. I. Individual and Total Rates of Decay of the Fission Products," U. S. Naval Radiological Defense Laboratory Report ADC-65, April 11, 1949 (U); Nucleonics 9, No. 5, C-2 (1951).

3. N. E. Ballou, "Radiochemical and Physical Chemical Properties of Products of a Deep Underwater Nuclear Detonation," Operation WIGWAM, Project 2.3, WT-1011, 17 April 1957. (SRD).

4. R. C. Bolles, and N. E. Ballou, "Calculated Activities and Abundances of U^{235} Fission Products," UCNRDL-456, 30 August 1956 (U); Nuclear Science and Engineering 5, 156-185 (1959).

5. L. E. Weaver, P. O. Strom, and P. A. Killeen, "Estimated Total Chain and Independent Fission Yield for Several Neutron-Induced Fission Processes," USNRDL-TR-633, 5 March 1963. (U).

6. J. L. Mackin, P. E. Zigman, P. O. Strom and M. J. Nuckolls, "Early Time Decay of Fission Product Mixtures. I. Scintillation Counter Measurements Following Thermal Neutron Fission of U^{235}," USNRDL-TR-359, 10 August 1959.(U).

7. P. E. Zigman, and J. Mackin, "Early Time Decay of Fission Product Mixtures. II. Gamma Energy Release and Ionization Rates Following Thermal Neutron Fission of U^{235}," USNRDL-TR-400, 11 February 1960. (U).

8. P. O. Strom, D. Sam, J. Mackin, and P. Zigman, "Early Time Decay of Fission Product Mixtures. IV. Gamma Ionization Chamber Measurements Following Thermal Neutron Fission of Pu^{239}," USNRDL-TR-528, 15 September 1961. (U).

9. O. I. Leipunsky, V. N. Saharov, V. I. Tereshchenko, "Gamma-Rays from Short-Lived Fission Products of U^{235} and Pu^{239}," Atomnaya Energiya 2, 278 (1957); Journal of Nuclear Energy 6, 170 (1957).

10. E. C. Freiling, "Radionuclide Fractionation in Bomb Debris," Science 133, 1991-1998 (1961).

11. E. C. Freiling, and N. E. Ballou, "Nature of Nuclear Debris in Sea-Water," Nature 195, 1283-1287 (1962).

12. E. R. Tompkins, and L. B. Werner, "Chemical, Physical, and Radiochemical Characteristics of the Contaminant," Operation CASTLE, Project 2.6a, WT-917, September 1955. (SRD).

13. T. Triffet, and P. D. LaRiviere, "Characterization of Fallout," Operation REDWING, Project 2.63, WT-1317, 15 March 1961. (SRD).

14. S. L. Whitcher, L. R. Bunney, R. R. Soule and R. A. daRoza, "Fallout Measurements by Aircraft and Rocket Sampling" (U), Operation HARDTACK, Project 2.8, WT-1625, September 29, 1961. (SRD).

15. W. R. Schell, and R. W. Caputi, "Fractionation of Fission Fragments from Underwater Nuclear Detonations Observed During Operation HARDTACK," USNRDL-TM-115, 16 October 1959. (C).

16. N. E. Ballou, and L. R. Bunney, "Radiochemical Measurements and Sampling Techniques. Nature and Distribution of Residual Contamination II," Operation JANGLE, Project 2.6c-2, WT-397 (bound with WT-373), 30 June 1952. (U).

17. E. C. Freiling, "Fractionation Correlations," Proceedings of Tripartite Symposium on Technical Status of Radiological Defense in the Fleets, USNRDL Reviews and Lectures 103, Vol. 1, p. 33, 10 June 1960; E. C. Freiling, "Fractionation. I. High Yield Surface Burst Correlations," USNRDL-TR-385, 29 October 1959. (U).

18. C. F. Miller, R. Cole, W. J. Heiman, "An Investigation of Synthetic Contaminants for an Atomic Burst in Deep Seawater," USNRDL-411, 6 August 1953 (SRD); C. F. Miller, Private communication.

19. C. F. Miller, R. Cole, W. J. Heiman, "Decontamination Reactions of Synthesized Fallout Debris for Nuclear Detonations. I. Nuclear Detonations in Sea Water," J. Colloid Science 13, 337-347 (1958).

20. L. R. Bunney, and N. E. Ballou, "An Estimate of the Predominant Chemical Species Resulting from a Shallow Underwater Burst of an Atomic Bomb," USNRDL-435, 5 February 1954. (U).

21. A. E. Greendale, and N. E. Ballou, "Physical State of Fission Product Elements Following Their Vaporization in Distilled Water and Sea Water," USNRDL-436, 16 February 1954. (U).

22. E. C. Freiling, "Physical State Studies," Proceedings of Tripartite Symposium on Technical Status of Radiological Defense in the Fleets, USNRDL Reviews and Lectures 103, Vol. 1, p. 47, 10 June 1960. (U).

23. E. C. Freiling, "Oxidation States of U, Np, and I in Nuclear Weapon Debris," Proceedings of Tripartite Symposium on Technical Status of Radiological Defense in the Fleets, USNRDL Reviews and Lectures 103, Vol. 1, p. 61, 10 June 1960. (U).

CONTROL OF ELEMENT VELOCITY DISTRIBUTIONS IN SONAR PROJECTOR ARRAYS

D. L. Carson, G. E. Martin, G. W. Benthien, J. S. Hickman
U. S. Navy Electronics Laboratory
San Diego 52, California

1. INTRODUCTION

In acoustic projector arrays, any particular directivity pattern and source level dictates a set of radiating head velocities v_1, $v_2, \ldots, v_p, \ldots, v_n$, or briefly $[v_p]$. To realize such a set of head velocities in an array of identical electromechanical radiators, it is a practical necessity that the head velocity of each radiator be proportional to its corresponding driving voltage or current, and further that the proportionally factor be identical for each radiator of the array. Thus, when the proportionality factor is known, the required set of driving voltages or currents, $[E_p]$ or $[I_p]$ can be computed from the required set of head velocities, $[v_p]$. When this condition prevails, we will say that voltage velocity control or current velocity control has been achieved.

It is an unfortunate fact of life that for an important class of arrays, namely those whose elements are closely packed and whose head dimensions are small compared to λ (the wave length of sound in water at operating frequency), the above factor of proportionality is not likely to be the same for each radiator of the array but changes from radiator to radiator across the array. This is a most unsatisfactory state of affairs, and the intent of this paper is to show why this is the case and to recommend corrective measures.

When the factor of proportionality varies from radiator to radiator, the set of driving voltages or currents fails to yield the anticipated set of head velocities, the radiation impedances computed for the anticipated set of head velocities are incorrect for the actual head velocities, and the resulting comedy of errors leads to chaos. The head velocities range in both magnitude and phase from very small values to sufficiently large values to cause structural damage, even at low drive levels. Expected acoustic source levels, patterns and bandwidths are not realized, and electric input impedance phase angles will be much larger than anticipated, sometimes exceeding $\pm$ 90 degrees.

For definiteness, we will center our attention on piezoelectric radiators, but the principles involved apply equally well to all

types.[1]

2. DIAGNOSIS

Regardless of how complicated a piezoelectric radiator of an array may be, it can be reduced to the four terminal Tee network shown in Fig. 1, which implies the equations

$$E_p = \frac{Z_{11}}{G}\left[\left(\mathfrak{z}_M - \frac{G^2}{Z_{11}}\right) + \mathfrak{z}_p\right] v_p \qquad (2.1)$$

and

$$I_p = \frac{1}{G}\left(\mathfrak{z}_M + \mathfrak{z}_p\right) v_p \qquad (2.2)$$

where E_p, I_p, v_p and z_p are, respectively, the driving voltage and current, the head velocity, and the radiation impedance associated with the typical or p^{th} radiator.

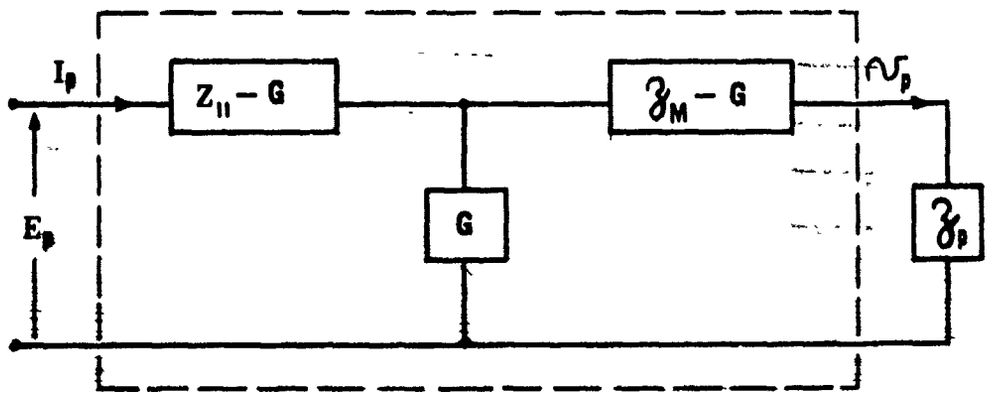


Fig. 1. Four terminal electromechanical equivalent circuit of an acoustic radiator.

On the other hand, Z_{11}, G, and z_M are, respectively, the blocked electrical impedance, the electromechanical coupling impedance, and the mechanical impedance, which, for a given radiator, depend only on frequency and are the same for all radiators of the array. Equations (2.1) and (2.2) show the proportionality factor for voltage drive and for current drive, respectively. By inspection, if, by accident or design,

$$\left| z_M - \frac{G^2}{z_{11}} \right| \gg |z_p| \qquad (2.3)$$

prevails over the frequency band of interest for a given set of head velocities, then, the proportionality factor in (2.1) would be approximately the same for all radiators of the array and voltage velocity control could be said to be inherent in the design. Similarly, if

$$|z_M| \gg |z_p|, \quad p = 1, \ldots, n, \qquad (2.4)$$

holds over the frequency band of interest, the factor of proportionality in (2.2) would be approximately the same for all radiators and we would say that current velocity control is inherent in the design.

This completes our diagnosis. If a given array suffers from poor voltage velocity control, then (2.3) does not hold. If it suffers from poor current velocity control, then (2.4) does not hold. This admittedly trivial observation is, nevertheless, the crux of the current widespread velocity control failures in modern low frequency high power sonar projector arrays.

3. CURES

The title of this section implies that our concern is with arrays suffering from poor velocity control. Remarks about other arrays will be reserved for a later section.

From section 2, we can be sure that poor voltage and/or current velocity control implies that the set of radiation impedances $[z_p]$, is not a set of identical numbers, and that (2.3) and/or (2.4) do not hold.

There are three approaches to a cure.

> **Cure A:** For a given set of velocities, $[v_p]$, use (2.1) or (2.2) to determine the required set of driving voltages, $[E_p]$ or currents $[I_p]$, which will yield $[v_p]$. Then develop a set of amplifiers, one for each radiator, which will supply $[E_p]$ or $[I_p]$ regardless of the set of radiation impedances $[z_p]$.
>
> **Cure B:** Attempt to achieve the equivalent of the condition (2.3) and/or (2.4).
>
> **Cure C:** Attempt to alter $[z_p]$ in such a way that all radiation impedances are alike.

Cure A is conceptually elementary, but may be extremely difficult to achieve. In this connection, we are studying the

possibilities of feedback from an accelerometer imbedded in each radiating head. The signal from the accelerometer on the p^{th} radiating head is fed back to the driving amplifier of the p^{th} radiator such that the phase and magnitude of v_p will remain constant relative to some reference.

Cure B has received our special attention to date and will be understood from the following. A ceramic LORAD array was suffering from poor voltage velocity control. We asked if some circuitry could be inserted between the radiator and its amplifier such that voltage velocity control could be achieved between v_p and some new voltage $\bar{E}_p$, other than E_p. Carson's work on variable reluctance radiators[1] suggested that a series inductor, L_c, should offer some promise. So a mathematical experiment was performed with a series L_c as shown in Fig. 2.

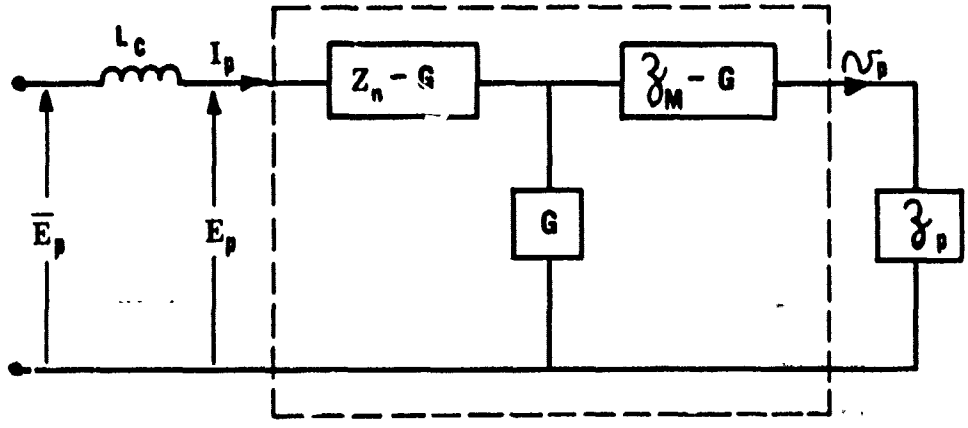


Fig. 2. Four terminal electromechanical equivalent circuit with a voltage velocity control inductor.

From Fig. 2, one can write

$$\bar{E}_p = \frac{(Z_{11} + j\omega L_c)}{G}\left[\left(\mathcal{Z}_M - \frac{G^2}{Z_{11} + j\omega L_c}\right) + \mathcal{Z}_p\right] v_p, \qquad (3.1)$$

or equivalently,

$$\left(\frac{G}{Z_{11} + j\omega L_c}\right)\bar{E}_p = \left[\left(\mathcal{Z}_M - \frac{G^2}{Z_{11} + j\omega L_c}\right) + \mathcal{Z}_p\right] v_p \qquad (3.2)$$

Thus (2.6) takes the form

$$\left(\frac{G}{Z_{11}+j\omega L_c}\right)\bar{E}_p = \left[\mu(\omega, L_c) + Z_p\right] v_p \qquad (3.3)$$

where the internal impedance, μ, is given by

$$\mu(\omega, L_c) = Z_M - \frac{G^2}{Z_{11}+j\omega L_c} . \qquad (3.4)$$

Evidently, if L_c can be selected so that

$$\left|\mu(\omega, L_c)\right| >> \left|Z_p\right| , \; p = 1, \ldots, n \qquad (3.5)$$

holds for every radiator of the array over the operational frequency band and for every required velocity distribution $[v_p]$, then voltage velocity control between $\bar{E}_p$ and v_p will have been achieved. We selected a mid band frequency, ω_o, and computed L_c to maximize $|\mu(\omega_o, L_c)|$. Then for ω_o, we computed the set of radiation impedance magnitudes, $[|z_p(\omega_o)|]$, for every radiator of the array for every set of required velocity distributions. We let $|z_p(\omega_o)|_{max}$ denote the maximum of all of these radiation impedance magnitudes and then observed that

$$\left|\mu(\omega_o, L_c(\omega_o))\right| >> \left|Z_p(\omega_o)\right|_{max} \qquad (3.6)$$

did indeed hold. $L_c(\omega_o)$ denotes the L_c selected to maximize $|\mu|$ at ω_o. Next we computed $|\mu(\omega, L_c(\omega_o))|$ and $|z_p(\omega)|_{max}$ over the operational frequency band and found, to our delight, that

$$\left|\mu(\omega, L_c(\omega_o))\right| >> \left|Z_p(\omega)\right|_{max} \qquad (3.7)$$

held sufficiently throughout the band for every set of velocity distributions required. We then procured n of these identical

inductors, each of value $L_c(\omega_o)$, inserted one such inductor in series with each radiator, and were able to verify experimentally that voltage velocity control had indeed been achieved. It was found that the mid band frequency choice for ω_o was not necessarily the best for balanced velocity control over a band, and that small adjustments in ω_o resulted in some improvement. Of course a new choice of ω_o implied a new choice of L_c to maximize $|\mu(\omega, L_c)|$.

We observe that a series inductor is no cure-all for velocity control. In particular, it would make no contribution whatever to current velocity control, since, by inspection, (2.2) holds for both Fig. 1 and Fig. 2 and L_c does not even appear in (2.2). Furthermore, it is conceivable that $L_c(\omega_o)$ could have failed to produce the inequality (2.10), in which case the series L_c scheme for voltage velocity control would have failed. On the other hand, some other additive electric circuitry might have achieved voltage velocity control, where L_c failed to do so.

For purposes of illustration, let us suppose that the above LORAD array had been known to be suffering from poor current velocity control. This was of no interest to us, since we were only interested in voltage drive. Nevertheless, the following scheme could have been considered. One could have examined mathematically the effects of some shunting reactor put across each radiator. Let us suppose, as shown in Fig. 3, that a shunting reactance, Z_c, has been added.

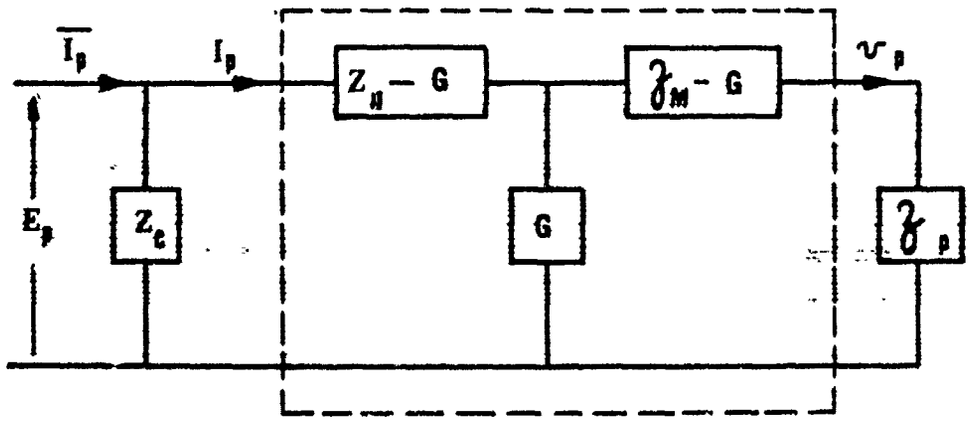


Fig. 3. Four terminal electromechanical equivalent circuit with a current velocity control reactance, Z_c.

Figure 3 implies the relations

$$E_p = \frac{z_{11}}{G}\left[\left(z_M - \frac{G^2}{z_{11}}\right) + z_p\right] v_p \qquad (3.8)$$

and

$$\bar{I}_p = \frac{(z_{11} + z_c)}{G z_c}\left[\left(z_M - \frac{G^2}{z_{11} + z_c}\right) + z_p\right] v_p \qquad (3.9)$$

First we note that since Z_c fails to appear in (3.8), a shunting reactance can have no effect on voltage velocity control. A comparison of the forms of (3.2) and (3.9) shows, by inspection, that current velocity control will be achieved in the latter if $Z_c = j\omega L_c$, where L_c is the same inductor which achieved voltage velocity control in (3.2). In (3.9), of course, L_c must be a shunting inductor. Evidently (3.9) can be expressed in the form

$$\left(\frac{j\omega L_c G}{z_{11} + j\omega L_c}\right)\bar{I}_p = \left[\mu(\omega, L_c) + z_p\right] v_p , \qquad (3.10)$$

where μ is defined by (3.4) and $L_c = L_c(\omega_o)$, as before.

Before discussing Cure C, it will be helpful to outline some known facts about radiation impedance. It can be shown that

$$f_p = -\sum_{a=1}^{n} z_{pa} v_a \qquad (3.11)$$

where f_p is the water borne force on the p^{th} radiating head resulting from the vibration of the radiating heads of the array. The coefficient, z_{pa}, is the negative of the ratio of the water borne force on the p^{th} head resulting from the vibration of the a^{th} head, to the head velocity of the a^{th} head. When $p = a$, we call z_{pa} the self impedance of the radiator. The z_{pa} are dependent on array geometry, but are independent of all head velocities. The equation, (3.11), can be extended to

$$f_p = -\left(\sum_a \zeta_{pa} \frac{v_a}{v_p}\right) v_p = -\zeta_p v_p, \qquad (3.12)$$

which defines the radiation impedance, z_p. A further breakdown of (3.12)gives

$$f_p = -\left(\zeta_{pp} + {}'\sum_a \zeta_{pa} \frac{v_a}{v_p}\right) v_p, \qquad (3.13)$$

where the prime to the left of Σ , denotes omission of a = p from the summation. From (3.12)and (3.13)we write

$$\zeta_p = \zeta_{pp} + {}'\sum_a \zeta_{pa} \frac{v_a}{v_p} = \zeta_{pp} + \bar{\zeta}_p \qquad (3.14)$$

which defines $\bar{z}_p$ as that part of z_p other than the self impedance, z_{pp}. Evidently z_p depends on the whole set of head velocities. Since we are supposing all radiators are identical, we have

$$\zeta_{11} = \zeta_{22} = \cdots = \zeta_{nn} \qquad (3.15)$$

Now let us examine the possibilities of achieving Cure C. One suggestion, which was communicated to the authors by Dr. A. G. Emslie of the Arthur D. Little Company, can be outlined as follows. From (3.12), we see that z_p is a function of the whole set of head velocities, $[v_p]$. One could ask if $[v_p]$ could be selected for a given array and frequency, such that all radiation impedances would equal the same complex number, ρ, to be determined. Thus we seek a set $[v_p]$, satisfying the n simultaneous linear homogeneous equations

$$\sum_a \zeta_{pa} \frac{v_a}{v_p} = \rho. \qquad (3.16)$$

This equation reduces to a polynomial of n^{th} degree ρ with coefficients which depend on the impedances, z_{pq}. Equation (3.16) can be rewritten to emphasize this polynomial in the form

$$P(\rho) = 0. \tag{3.17}$$

Now (3.17) will certainly have one solution and in general will have n solutions, $\rho_1, \ldots, \rho_n$. Corresponding to each ρ_i there will be a velocity distribution, $[v_p]_i$, such that

$$\sum_a \zeta_{pa} \frac{v_a}{v_p} = \rho_i \,. \tag{3.18}$$

The set, $[v_p]_i$, and the solution, ρ_i, could be substituted in (2.1) or (2.2) and the appropriate driving voltages, $[E_p]_i$, or currents, $[I_p]_i$, could be determined from

$$E_p = \frac{z_{11}}{G}\left[\left(\zeta_M - \frac{G^2}{z_{11}}\right) + \rho_i\right] v_p \tag{3.19}$$

or

$$I_p = \frac{1}{G}\left(\zeta_M + \rho_i\right) v_p \,. \tag{3.20}$$

This scheme is theoretically possible, but has some serious practical difficulties. Ordinarily, the set of directivity patterns required of an array is given as part of the design specifications. Usually there will be a pattern for each direction of transmission, and these different patterns are often achieved by critical adjustment of the relative phases and magnitudes of the head velocities. Thus, complete freedom of selection of the set of head velocities is usually mandatory. It is evident that the velocity control scheme under discussion limits our velocities to the particular derived sets, $[v_p]_i$, corresponding to $\rho_1, \ldots, \rho_n$, and none of these sets were selectable by us. It would be fortuitous if any of the derived sets

of head velocities coincided with any of the sets of head velocities dictated by the design specifications. In other words, this scheme of velocity control precludes phase steering and in general allows no freedom of choice in the selection of velocity distributions. Moreover, it requires that we be able to supply, corresponding to ρ_i, the required set of voltages, $[E_p]_i$, or currents $[I_p]_i$, to achieve the derived set of velocities $[v_p]_i$. One may well ask, if we have this essentially unlimited ability to provide driving voltages or currents, why limit ourselves in our choice of velocity distributions? Why not use Cure A and be done with it? In spite of these disadvantages, the idea of identical radiation impedances is intriguing, and if it should happen that one of the derived velocity distributions coincided with one dictated by the design specifications, this knowledge might be very useful.

There is another way to realize Cure C which also is theoretically possible, but quite unacceptable. It can be shown, relative to (3.14), that

$$|\zeta_{pp}| >> |\bar{\zeta}_p|, \quad p = 1, \ldots, m \qquad (3.21)$$

will hold if the radiators are sufficiently removed from each other, or if the radiating heads have dimensions which are a large fraction of λ (the wave length of sound in water). Equations (3.21) and (3.14) imply, to first order, that

$$\zeta_p = \zeta_{pp}, \qquad (3.22)$$

or, by (3.15), that all radiation impedances are alike and equal to the self impedance of a single radiator. It is well known, that the self impedance of a radiator whose head dimensions are a small fraction of λ, is largely reactive with only a small resistive component. These undesirable characgeristics disappear if the radiating head dimensions are a large fraction of λ. Thus since the array under consideration has poor velocity control, the radiation impedances must differ over the array, so (3.22) could not hold, and we must conclude that our sick array must have heads which are a small fraction of λ, and the array must be tightly packed. Thus, to be sure, one could achieve good velocity control by simply separating the radiators sufficiently from each other. One might say that separating radiators alters array geometry and thus violates the rules of our game. But more to the point, the separated radiators will see only their self impedance, which will be highly reactive with small resistive component, since the radiating heads are small.

Finally, then, poor velocity control can be cured by spreading the radiators apart, but array efficiency, bandwidths, and realizable source level will all suffer an unacceptable degradation.

In summary, we can say that Cure A is excellent, provided we can acquire driving amplifiers which will supply exactly the required set of voltages or currents specified by (2.1) or (2.2) to force the radiators to have the required set of velocities. Of course, in this event, even if all head velocities are identical, the set of voltages or currents will vary in magnitude and phase from radiator to radiator following the changing radiation impedances. Cure B has been used successfully in a large LORAD variable reluctance array and in the large LORAD ceramic array now operating with good velocity control aboard the submarine, BAYA. This Cure is relatively inexpensive, and perhaps offers the best overall solution to velocity control. The approach to Cure C which attempts to force all radiation impedances to be alike may offer some promise in certain special cases, but the idea of loosely packing the radiators to force uniformity of radiation impedance appears to offer no promise whatever.

We have devoted our entire discussion to arrays which exist and which exhibit poor velocity control. Of course, when starting a new design, one should consider carefully the possibility of using radiators with large head dimensions to avoid the need for a scheme of velocity control. Some obvious problems at low frequency with large radiators are their large weight with attendant handling difficulties, as well as the fact that it is not easy to avoid spurious modes of vibration in large vibratory structures. Then, too, the larger radiating heads tend to limit beam pattern formation and control of phase steering.

4. DISCUSSION

To keep the central ideas of our paper before us, we have consciously delayed remarks which would have a diverting influence. Now seems an appropriate time to pick up these threads.

In World War II ultrasonic arrays, the dimensions of radiator heads were set by the size of available crystals on magnetostrictive rods. Fortuitously, these dimensions were rarely less than an inch, which at ultrasonic frequencies was a large fraction of λ. Thus, in such arrays, velocity control was never a problem, since (3.21) held, albeit accidentally. By adequate but crude techniques, as compared with our current methods, a single radiation impedance, z, was computed, and every radiator was assumed to have this same impedance. Then, by filter techniques, a projector sensitivity bandwidth was computed from Fig. 1, using z as the radiation impedance, and this was taken as the theoretical bandwidth of the array. Ultrasonic arrays designed in this manner exhibited calibrated sensitivity bandwidths which compared well with theoretical predictions. As operational frequencies dropped and array sizes increased, radiator head dimensions failed to increase proportionally, with the result head dimensions soon measured a small fraction of λ. It was realized that, at these lower frequencies, radiation impedances would

differ from element to element, and, indeed, individual radiation impedances were computed for each radiator on the faulty premise that good velocity control was inherent in all arrays. The common practice was to determine some mean value, $\bar{z}$, of the whole set of computed radiation impedances. This value was inserted for z_p in Fig. 1, and as before, filter techniques were used to compute a theoretical projector sensitivity bandwidth which again was thought to be the operational bandwidth of the array. Of course, this bandwidth never came close to being realized in practice, because, in the chaos of poor velocity control, sensitivity bandwidth in the above sense lost all meaning.

The lengthy discussion in the last paragraph was intended to set the stage for a more realistic examination of bandwidth. We have seen that good velocity control will be realized only over that band for which (3.5) holds. Such a band will be called, velocity control bandwidth, for identification. Our experience to date indicates that the velocity control bandwidth tends to be narrower than the sensitivity bandwidth discussed in the preceding paragraph.

The crux of the matter is that any bandwidth, however computed, which exceeds the velocity control bandwidth has no operational significance, since the actual array performance completely deteriorates outside the velocity control band.

It might appear that velocity control based on the condition (3.5) would imply a marked reduction in voltage sensitivity, $|v_p/E_p|$ in (3.3) or current sensitivity, $|v_p/I_p|$ in (3.10). This is not the case, however, for, to first order, $|\mu|$ is maximized by that choice of Z_c which maximizes $|1/(Z_{11} + Z_c)|$. Inspection of (3.3) and (3.10) shows that this same choice of Z_c maximizes the effective driving voltage and current, $|G\bar{E}_p/(Z_{11} + Z_c)|$ and $|Z_0 G\bar{I}_p/(Z_{11} + Z_c)|$. So the magnitudes of the internal impedance, $|\mu|$, as well as the effective driving voltage and current all increase with $|1/(Z_{11} + Z_c)|$.

Velocity control via Cure B has no effect on efficiency so long as the additive control impedances are purely reactive.

Heretofore, we have concentrated entirely on the theory of velocity control. However, there are some technical aspects which must be appreciated if the full potentialities of the theory are to be realized. For example, let us reconsider the LORAD array. It utilizes modular drive where each radiator has its own driving amplifier. A typical amplifier, radiator, and voltage velocity control inductor, L_c, are shown in Fig. 4, where the input signal voltage, E_s, may or may not be the same for all radiators.

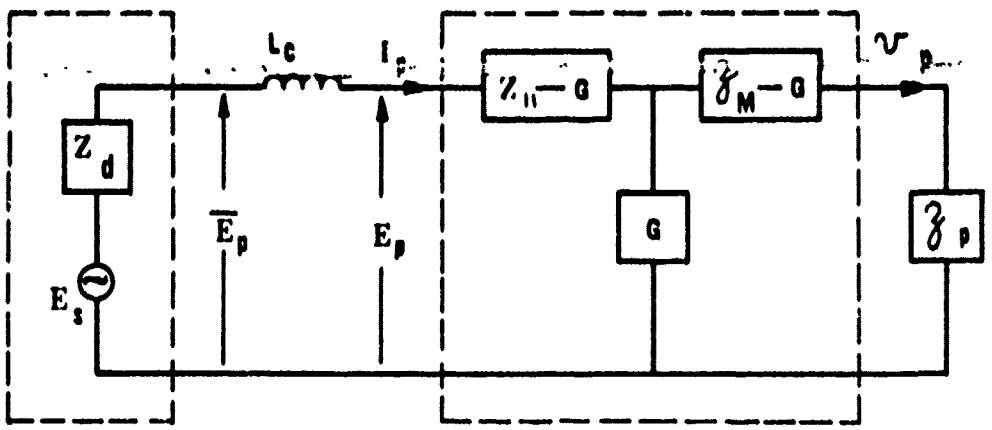


Fig. 4. Four terminal electromechanical equivalent circuit with a voltage velocity control inductor, L_c, and driving amplifier.

Observe that earlier we achieved voltage velocity control between $\bar{E}_p$ and v_p by a careful selection of L_c. It is clear that we will not have proper voltage velocity control between E_s and v_p unless $E_s = \bar{E}_p$, or equivalently, the internal impedance of the driver, Z_d, is zero. Since such zero impedance drivers can only be approximated in practice, the LORAD drivers were designed for minimum internal impedance to disturb the value of L_c as little as possible.

In Fig. 4, as well as in general, a distinction must be made between the notion of the voltage velocity control inductor, L_c, and a series tuning inductor, say L_T. Normally the latter would be selected purely for impedance matching, and typically would be selected to satisfy the relation

$$\frac{E_p(\omega_o)}{I_p(\omega_o)} + j\omega_o L_T = \text{a real number} \qquad (4.1)$$

which obviously has nothing to do with the criterion used for the selection of L_c. In fact, in Fig. 4, if impedance matching is required, some type of shunting reactance must be used, perhaps a shunting inductor, L_T, as shown in Fig. 5, and defined by the relation

$$\frac{\bar{I}_p(\omega_o)}{\bar{E}_p(\omega_o)} = \frac{1}{j\omega_o \bar{L}_T} + \frac{1}{j\omega_o L_c + \frac{\bar{E}_p(\omega_o)}{\bar{I}_p(\omega_o)}} = \text{a real number} \qquad (4.2)$$

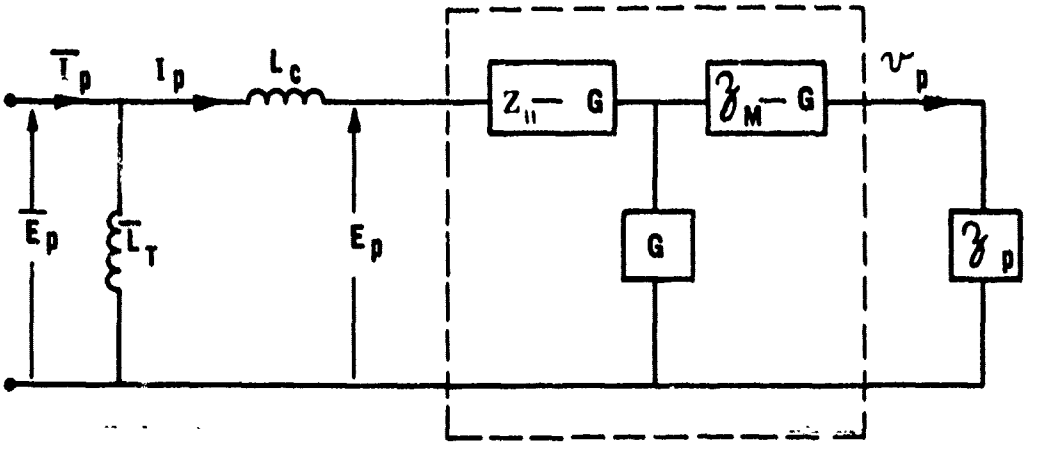


Fig. 5. Four terminal electromechanical equivalent circuit with voltage velocity control inductor, L_c, and impedance matching inductor $\bar{L}_T$.

If this shunting inductor would achieve impedance matching, it would do so without altering in any way the voltage velocity control between $\bar{E}_p$ and v_p.

Of course, similar remarks can be made relative to current velocity control. Evidently, for impedance matching in this case, a series tuning reactance would be added to the left of Z_c in Fig. 3 in such a manner that $(\bar{I}_p/v_p)$ is unaltered by the addition.

We have confined our attention to piezoelectric radiators, but if the principles discussed above are understood, they can easily be applied to magnetic and other types of radiators by a designer with the overall theoretical understanding of his subject he should have.

Some further examples of circuitry for velocity control using Cure B, are pertinent. Suppose an array of n radiators is driven in parallel from a single amplifier. A suitable arrangement for voltage velocity control is shown in Fig. 6, where L_p is an impedance matching inductor.

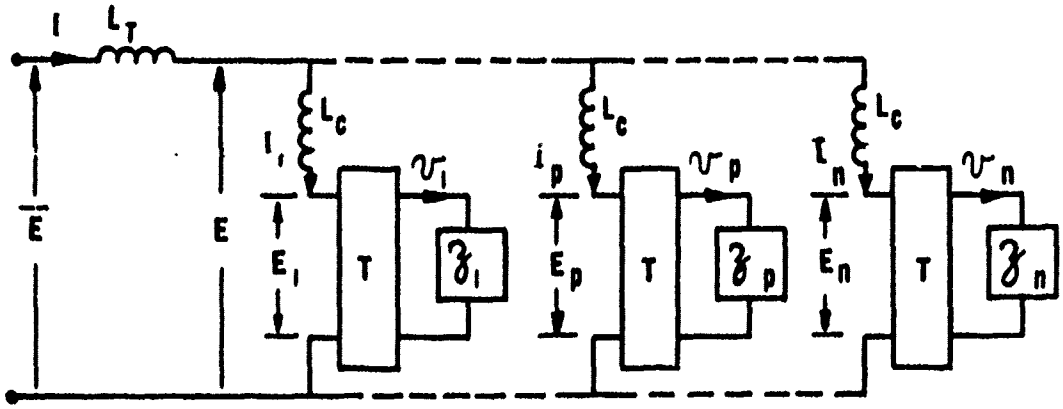


Fig. 6. Electroacoustic radiators driven in parallel with voltage velocity control and impedance matching inductor.

Figure 6 will assure that all head velocities are identical if L_c is selected for velocity control between E and v_p. Of course L_T has no effect on voltage velocity control, since the velocity control in this case is between E and the head velocities. A shunting reactance could have been used instead of L_T for impedance matching. At the risk of reduncancy, we point out that if the control inductors, L_c, are shorted out in Fig. 6 we have destroyed our velocity control.

Suppose an array of n radiators is driven in series. A suitable arrangement for current velocity control is shown in Fig. 7, where all head velocities will be identical if L_c is selected appropriately and current velocity control between I and v_p will exist.

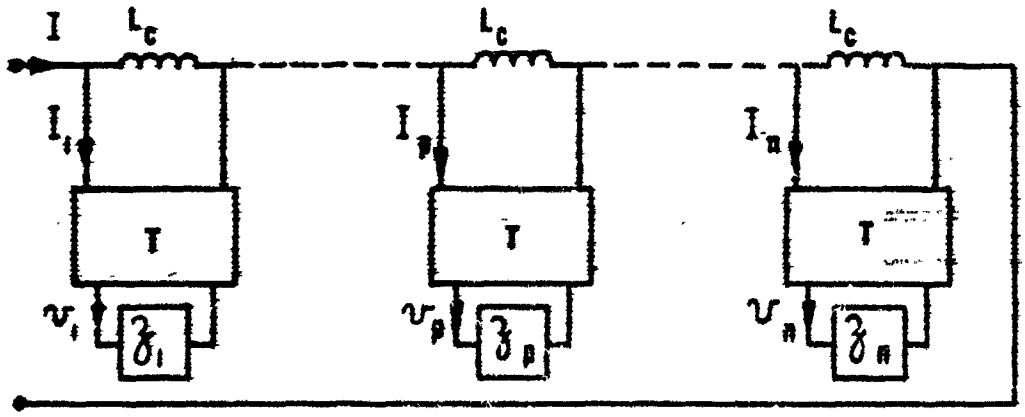


Fig. 7. Electroacoustic radiators driven in series with current velocity control.

Any impedance matching circuitry would be added to the left of 1 in Fig. 7. Again, the removal of L_c will destroy the velocity control.

In trying to emphasize the basic concepts of velocity control, we have purposely played down the complexities inherent in configurations such as Fig. 1. For a single ceramic LORAD radiator, for example, the four terminal network shown in Fig. 1 represents the configurations shown in Figures 8a and 8b where the latter is a greatly simplified electromechanical equivalent circuit of the former.

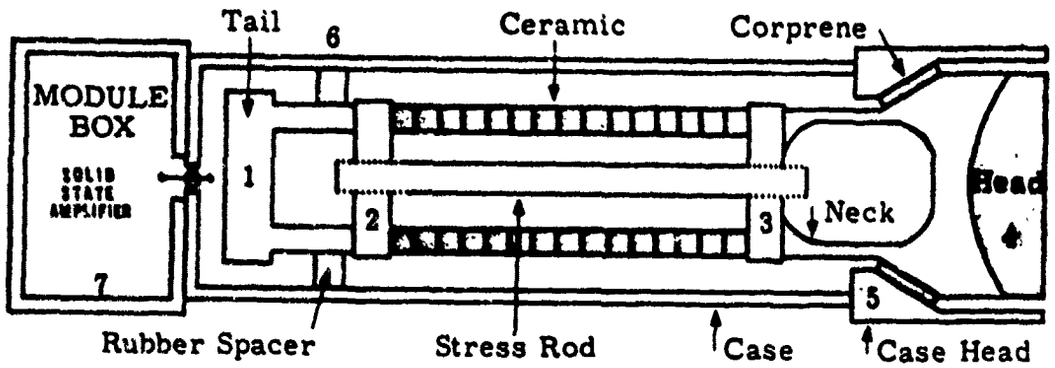


Fig. 8a. Cross section of LORAD ceramic radiator.

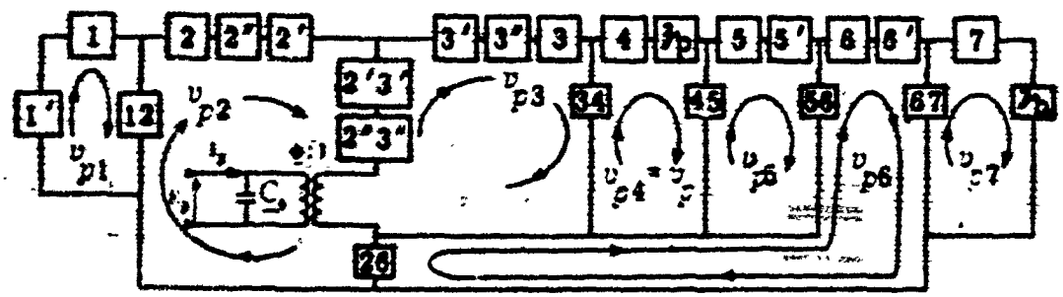


Fig. 8b. Simplified electromechanical equivalent circuit for the radiator shown in Fig. 8a.

All of the impedance boxes in Fig. 6b are rational functions of circular and hyperbolic trigonometric functions with complex arguments wherein frequency, and temperature, as well as stress and electric bias appear as parameters. The radiation impedance, z_p, shown in Fig. 8b is typically an infinite series derived from boundary value considerations, in which appears the entire set of head velocities $[v_p]$, and which sometimes converges so slowly that several hundred terms of the series must be evaluated for every frequency and for every radiator. Since such radiators interact on each other through mutual radiation impedance, it is not uncommon for the array designer to be confronted with the task of solving many simultaneous equations- some 3600 such simultaneous equations were involved in the LORAD design. Such equations, of course, do not appear out of nowhere, but must be laboriously built up from non-linear, dissipative, elastic equations of state with superimposed electric and thermodynamic fields.

Velocity control is not something apart from all this, but rather is a small but vital portion of a much larger design problem. Consequently the only general recipe we can give on velocity control is to assert that after one has mastered the larger overall problem, he is then in a position to attack velocity control per se. There seems to be no easier way.

ACKNOWLEDGMENTS

The authors wish to express their thanks to J. Elliot for his suggestions and constructive criticisms, and to Mrs. E. T. Schwarz, G. R. Anthony and H. B. Peters for their assistance relative to manuscript preparation.

REFERENCES

(1) D. L. Carson: Diagnosis and Cure of Erratic Velocity Distributions in Sonar Projector Arrays, Journal of the Acoustical Society of America, Vol. 34, No. 9, Part 1, pp 1191 - 1196, September, 1962.

SOLID STATE SPECTROSCOPY OF RARE EARTH IONS

R. A. Buchanan, J. Murphy and H. H. Caspers
U. S. Naval Ordnance Laboratory
Corona, California

INTRODUCTION

The rare earth elements (type 4f) are those members of the periodic table with atomic numbers 57 through 71. They are characterized by an electronic structure possessing partially filled 4f orbitals ($n = 4$, $\ell = 3$). The rare earth elements are then generated primarily by filling of the 4f shell; lanthanum, the first of the series, has no f electrons; and lutetium, the last of the series, has fourteen 4f electrons. The energy levels of each atom or ion are formed by the different configurations of its f electrons.

The 4f electrons apparently lie mostly within the orbitals formed by the $5s^2$, $5p^6$, and $6s^2$ electrons. As a result, the levels formed by the 4f electronic configurations are shielded by the outer electrons. When the rare earth ion is incorporated into a crystal, the perturbation due to the crystalline field on the 4f orbitals is small compared with the coulomb interaction between electrons and hence, the positions of the free ion energy levels are essentially independent of the host lattice. Because of the shielding, many of the 4f energy levels remain very narrow when the ion is incorporated into a crystal, and radiative transitions between these levels are quite common. These are the features which make the rare earths of interest as optical devices.

The 4f electronic energy levels of interest in this discussion are the trivalent rare earths which are now very well understood, due mostly to the work of Professor Dieke and his group at Johns Hopkins University. Figure 1 shows a table of the energy levels which Professor Dieke has compiled. Across the bottom of the figure are listed the rare earth elements along with the spectroscopic notation for their ground state. Increasing upward along the ordinate is the energy listed in units of cm^{-1}. These level diagrams are essentially complete.



Fig. 1 Energy level diagram of trivalent rare earth ions

We have taken the liberty of including on the diagram the transitions which have been reported to give rise to stimulated coherent emission. We have said that these levels are essentially independent of the host lattice into which the ion may be incorporated, and since these levels have been determined, one may ask "what remains to be learned?" The answer to this question lies mainly in the effects which are produced by incorporating the ion into the crystal lattice. Although the effect of the crystalline field on the overall energy level diagram is small and is indicated in the figure by the width of the levels, it is not negligible. The crystalline field is the predominant perturbation which determines the widths of spectral lines, the lifetimes of states and the transition probabilities between states. It is due to these effects that certain ions will produce stimulated coherent emission in one type of host lattice, but not in another.

Crystalline Field Effects

In the free ion, transitions which take place within the states of a given configuration must be either magnetic dipole or electric quadrupole in nature. This is so because all states of the same electron configuration have the same parity. Electric dipole transitions can only take place between states of opposite parity.

Suppose, now, that we consider an electron in a situation in which its potential energy function shows complete spherical symmetry with respect to the origin. Spherical symmetry implies conservation of angular momentum and conservation of parity, the two being intimately connected. There is a distinction, however, since one could have conservation of parity without conservation of angular momentum but not vice-versa. Now suppose the symmetry of the potential in which the electron finds itself is reduced to that which exists in a crystal. As has just been observed, the removal of symmetry requires a breakdown of some conservation law. If the crystalline potential retains inversion symmetry at the site of the ion with which the electron is associated, then only the conservation of angular momentum law is broken, and the wave functions which describe the state of the electron must consist of a linear combination of angular momentum functions. These functions must, however, all have the same parity because we have supposed inversion to be a symmetry operation. If, in addition, we now remove the inversion symmetry by either allowing the ion of interest to vibrate away from the inversion center or place it in a crystal in a non-inversion symmetry site, then parity is no longer conserved. The wave functions now contain both even and odd parity components. Most of the transitions which occur in the 4f-4f region of the spectra of rare earth ions have been found experimentally to be electric dipole in nature (some magnetic dipole transitions have been reported). These are referred to as "forced electric dipole" because they are allowed only through the parity mixing interactions discussed above. The nomenclature 4f in these cases really indicates that the level is mostly

4f. The nearest configuration is the 5d which is probably the stronger mixture with 4f. The transitions take place between the odd parity part of one state to the even parity part of another. It is possible that in some states no parity mixing is allowed because of selection rules imposed by the particular symmetry. In such cases, one may observe magnetic dipole transitions.

The strength of the interaction of the electric field with electrons in the ion is always less than the coulombinteraction between the electrons themselves; however, it may or may not be greater than the interaction between the spin and orbital magnetic moments of the electrons. In the case of the transition metal ions, crystalline field effects are much larger than spin-orbit effects, whereas the reverse is true for rare earth ions. The reduced crystalline field for rare earths has been attributed to the shielding of the electrons in the 4f shell by those in the outer filled 5s and 5p shells. The transition metals have no such protection because the incomplete shell in this case is the 3d shell which is the outermost shell of the ion.

The spectra of rare earth ions in crystals corresponding to transitions within the 4f energy levels, generally consist of sharp lines. The transition metal ion spectra show some sharp lines but, in general, the lines are quite broad. The reason for this difference is that the 4f electrons lie well within the ion and so their wave functions do not overlap the electron wave function on the ligands. The transition metal electrons, however, do overlap the ligand wave functions and this gives rise to the broadening. It is possible that in some states the overlap is still negligible even for the transition metals and this would account for the sharp lines observed. The idea of shielding of the crystalline electric field by outer electrons has been under attack recently by G. Burns[1]. He argues that outer electrons do not shield the 4f electrons in the rare earth by more than 10%. According to Burns, the reduction of the crystalline field effect in comparison with the spin orbit interaction is due to the following three effects:

a) The rare earths are larger than transition metal ions due mainly to the outer 5s and 5p filled shells. The ligands must then be farther away from the ion 4f shell. The crystalline field parameters (in the point charge model) depend on the separation of the paramagnetic ion and the ligands as

$$A_n \propto \frac{1}{R^{n+1}}$$

where A_n is a coefficient in the expansion of the crystalline field and R is the separation of the rare earth ion from a ligand.

b) The expectation values of r^n where r denotes the radial coordinate for a 4f electron are roughly 2 times smaller than for a 3d electron.

c) The spin-orbit coupling parameters in the rare earths are about twice as large as they are in the transition metals.

These effects all act to make the crystalline field interaction smaller than the spin-orbit interaction in the rare earth ions. In addition to the effects of R and r^n, the overall splitting of a free ion level depends upon the magnitude of the angular momentum.

The modern approach to the study of crystalline field effects involves absorption and fluorescence spectroscopy as well as electron paramagnetic resonance. EPR is usually studied in the ground state. However, some work has been done on EPR in excited states of rare earth ions[2]. The amount of splitting of an energy level when the ion is placed in a magnetic field is strongly dependent upon the crystalline field. EPR experiments are designed to measure this splitting and hence measure the crystalline field effects. On the other hand, the separation of the electronic energy levels due to crystalline field splitting is observed directly in optical absorption and fluorescence experiments. Provided a sufficient number of energy level separations can be obtained, the strength of the crystalline field can be determined. From the crystalline field, so determined, one can calculate the splitting which would be expected due to the addition of an external magnetic field. This splitting must agree with that determined by EPR techniques.

The agreement between the various experiments is usually quite good; however, one must consider all results in the light of the approximations made. First of all, the model which is chosen to represent the field is that of a system of point charges or dipoles surrounding the paramagnetic ion. This, of course, assumes that the electrons which give rise to the paramagnetism do not overlap the ligand electrons at all, i.e., they play no part in the binding. This approach is obviously wrong in the case of the transition metal ions where the unfilled 3d shell is the outermost shell. In this case, a 3d electron and two 4s electrons are used to form the bonding orbitals which hold the ion to the crystal.

The rare earth case is less obvious since the electrons spend most of their time within the 5s and 5p orbits and so, although they may not be electrically shielded to a great extent [1], they nevertheless are mechanically shielded from overlap with the ligands.

Attempts have been made[3] to include overlap of the paramagnetic ion orbitals with those of the ligand. The results indicate that the strength of the crystalline field depends upon the transformation properties of the electron wave function, i.e., if the

crystalline potential is expanded in spherical harmonics Y_ℓ^m

$$V = \sum_{\ell,m}' A_\ell^m r^\ell Y_\ell^m (\theta, \phi)$$

then the coefficients A_ℓ^m are not constant even for different states arising out of the same configuration. This situation is too difficult to treat by the present methods and much more work needs to be done in this area.

In the spectra of rare earth ions in crystals, it is often found that the number of transitions observed by far exceeds the number expected on the basis of a static crystalline field. The problem of trivalent ions in divalent lattice sites is complicated by the requirement that the extra positive charge must be compensated in order to maintain the overall neutrality of the crystal. The various charge compensating mechanisms will be discussed in detail in the next section.

In crystals where charge compensation is not present, we must look for another mechanism to explain the number of lines observed. Until now, we have been discussing the interaction of the 4f electrons with a static crystalline field. The host lattice is, itself, a system with energy levels. The energy states are the normal modes of vibration of the lattice. Since an interaction exists between the 4f electrons and the lattice, the two systems are coupled together. The energy levels of the combined system are then superimposed. The absolute values of the energies will, of course, shift slightly due to the interaction energy. It is not at all surprising then that more transitions will be allowed in the coupled electronic vibrational system than in the electronic system alone. Using group theoretical techniques, one can determine the types of lattice vibrations which are allowed by symmetry to couple to any given electronic level. The states of the coupled system between which transitions are allowed are also determined group theoretically. On the basis of these selection rules, it is sometimes possible to identify lattice vibrations from an absorption spectrum of the rare earth ion in the crystal.

A feature of the dynamic interaction of the rare earth ion and the lattice is the breakdown of static field selection rules. In the static field or uncoupled system, one determines the number of allowed transitions. In the coupled system, the electronic levels are mixed together so that the new states may have components to which transitions from the ground state are allowed.

The energy states of the lattice are in general, much broader than the energy states of the 4f electrons. The states of the coupled system, therefore, become broadened by the interaction with

the lattice. If the electronic levels are separated by energies approaching those of the lattice vibrations, an electron in one state can relax to another state and at the same time excite a lattice vibration. The process is known as a radiationless transition and is very effective in reducing the lifetime of an electron in excited states. By such processes, excited electrons may filter down through the ladder of closely spaced energy levels and find themselves in a level which is far removed from any electronic level. When this happens, the only way the electron can lose energy is to radiate a photon.

Charge Compensation

The problem of interpreting and, hence, understanding spectra of trivalent rare earth ions in divalent host lattices is sometimes complicated by various charge compensation mechanisms. In CaF_2, for example, there are at least five ways which have been proposed to compensate the extra positive charge of the trivalent rare earth occupying a divalent calcium site. These are as follows: An extra negative charge may lie somewhere in the crystal but sufficiently removed such that the trivalent rare earth has essentially the cubic site symmetry of the calcium ion. An interstitial fluoride ion may lie adjacent to the rare earth. Two site symmetries are possible for this arrangement; O^{2-} may substitute for an F^-, and finally, an alkali ion such as Na^+ may substitute for an adjacent Ca^{2+}. Each charge compensation arrangement will produce a different crystalline field at the site of the rare earth ion and each such field will produce its own characteristic spectrum. A trivalent rare earth ion may therefore have several distinct spectra when placed in CaF_2. The charge compensation mechanisms present in a given sample depend upon crystal growth conditions.

Figure 2 shows the spectrum of Tm^{3+} in CaF_2. It is determined from these spectra that there are several charge compensating mechanisms operating in this particular crystal. Work is now proceeding to separate out the various spectral lines and assign them to their appropriate charge compensation mechanisms.

Spectra of Ce^{3+}

The $4f^1$ electronic configuration of Ce^{3+} gives rise by means of the spin-orbit interaction to a $^2F_{7/2}$ and a $^2F_{5/2}$ level in the free ion. These levels have been determined from free-ion spectroscopy[4] to be separated by 2253 cm^{-1}. The spectrum for Ce^{3+} in any crystal would thus be expected to appear in the vicinity of 2253 cm^{-1} or about 4.5 microns.

Several studies[5,6] indicate that the maximum separation observed in a given Stark split free ion level is about 250 cm^{-1} or less for Pr^{3+} and Nd^{3+} in $LaCl_3$. Maximum Stark splittings of about 250 cm^{-1} are common for most rare earths in $LaCl_3$.

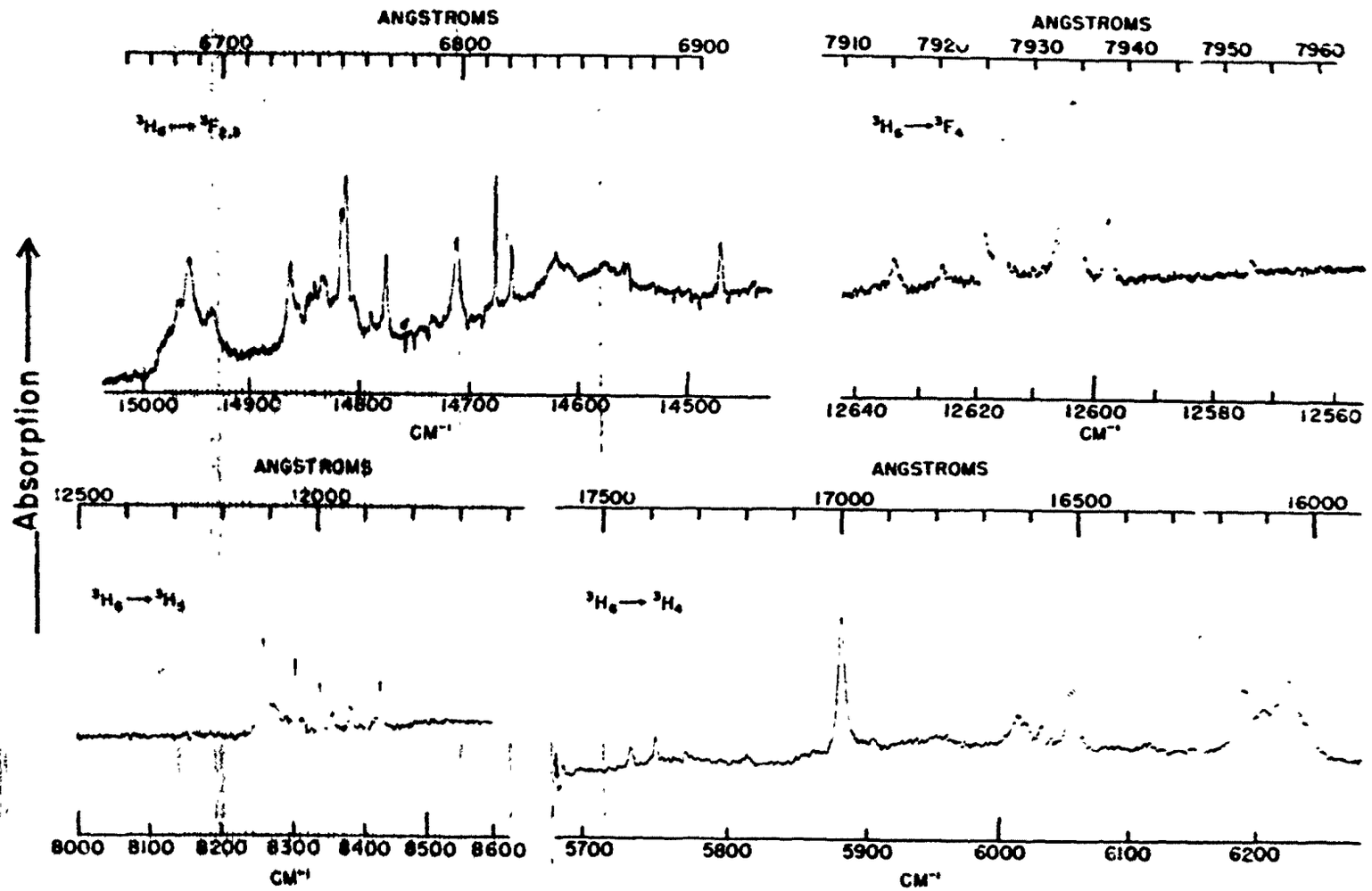


Fig. 2 Tm^{3+} in CaF_2

The site symmetry for La in $LaCl_3$ and LaF_3 is C_{3h} and C_{2v} respectively. It is assumed that cerium enters these hosts by substituting for a lanthanum ion and, hence, should be subject to a crystalline field of the appropriate symmetry. Electron paramagnetic resonance data [7] has confirmed that the site symmetry for Ce^{3+} in $LaCl_3$ is, indeed, C_{3h}. In LaF_3 EPR data [8] indicate that no symmetry at all exists at the Ce^{3+} site. This is, indeed, a surprising result. Furthermore, the EPR data [8] reveal that while Gd^{3+} has the La site symmetry C_{2v}, most other rare earth ions have no symmetry when substituted in LaF_3.

The cerium site symmetry determines the selection rules governing transitions between various crystalline field split levels [9]. Two electric dipole transitions from the ground state to the upper multiplet states in Ce^{3+} in $LaCl_3$ are allowed. Four such transitions can occur for Ce^{3+} in LaF_3. From these considerations alone one would predict that the electric dipole absorption spectrum of Ce^{3+} in $LaCl_3$ and LaF_3 near absolute zero temperature should consist of 2 and 4 absorption lines respectively. These transitions are indicated in Fig. 3. If, in addition, magnetic dipole transitions are important as well as the electric dipole transitions, then each spectrum should consist of 4 lines.

The spectra of Ce^{3+} in LaF_3 and $LaCl_3$ near 4.2°K are shown in Fig. 4. These two spectra exhibit many similarities. The very weak lines at low cm^{-1} become more intense at liquid nitrogen temperature and therefore must be associated with transitions which originate on excited energy levels. Intense and sharp lines are observed between 2100 and 2200 cm^{-1}. Two intense and broad lines containing complex structure are observed between 2500 and 3500 cm^{-1}. In between the sharp line region and the broad line region lies a region which contains many weak broad lines. Clearly, there are too many lines to be accounted for by pure electronic transitions. The sharp lines between 2100 and 2200 cm^{-1} are most likely pure electronic transitions. The weak broad lines in the central region are vibronic lines, i.e., lattice vibration frequencies coupled to the sharp electronic lines between 2100 and 2200 cm^{-1}. These vibronic lines will be discussed shortly.

The intense broad lines in the 2900 to 3500 cm^{-1} region are more difficult to understand. If these are indeed electronic transitions, they may be broadened and intensified by mixing with lattice vibration frequencies coupled to the lower electronic levels. Such effects have been noted previously [10]. The overall multiplet splitting is about 700 cm^{-1} in LaF_3 and about 1200 cm^{-1} in $LaCl_3$. This is much larger than the 250 cm^{-1} observed for Pr^{3+} and Nd^{3+} in $LaCl_3$. A crystalline field splitting of 1000 cm^{-1} has been reported [11] for Ce^{3+} and Yb^{3+} in CaF_2. Although the exact assignments of the absorption peaks of Ce^{3+} in $LaCl_3$ and LaF_3 are not known at this time, it appears probable that the crystalline field splitting of the Ce^{3+} multiplet levels is anomalously large.

Samples: Crystalline $LaCl_3$ and LaF_3 Doped with Ce^{3+}

Atomic Electron Structure: La - (Xenon) $5d^1\ 6s^2$

Ce - (Xenon) $4f^1\ 5d^1\ 6s^2$

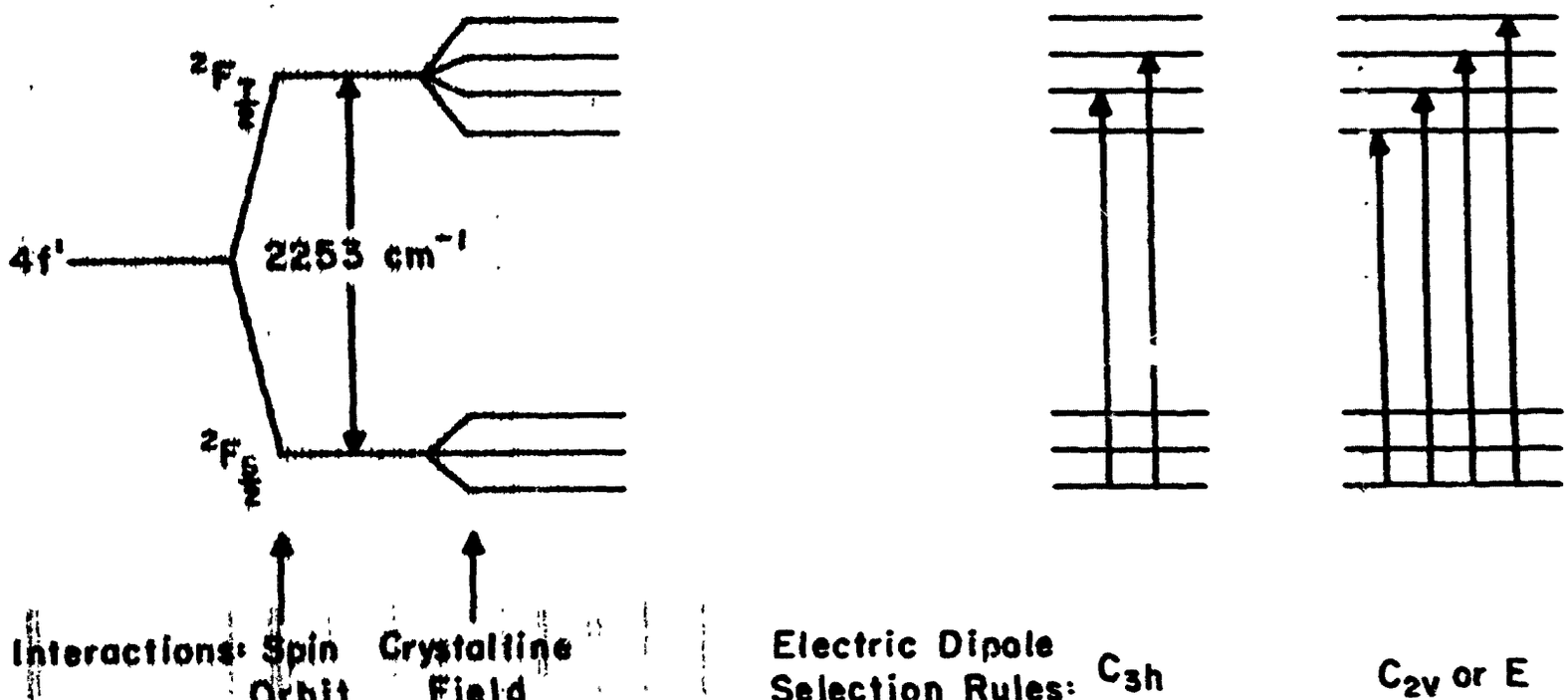


Fig. 3 Ce^{3+} Transitions in $LaCl_3$ and LaF_3

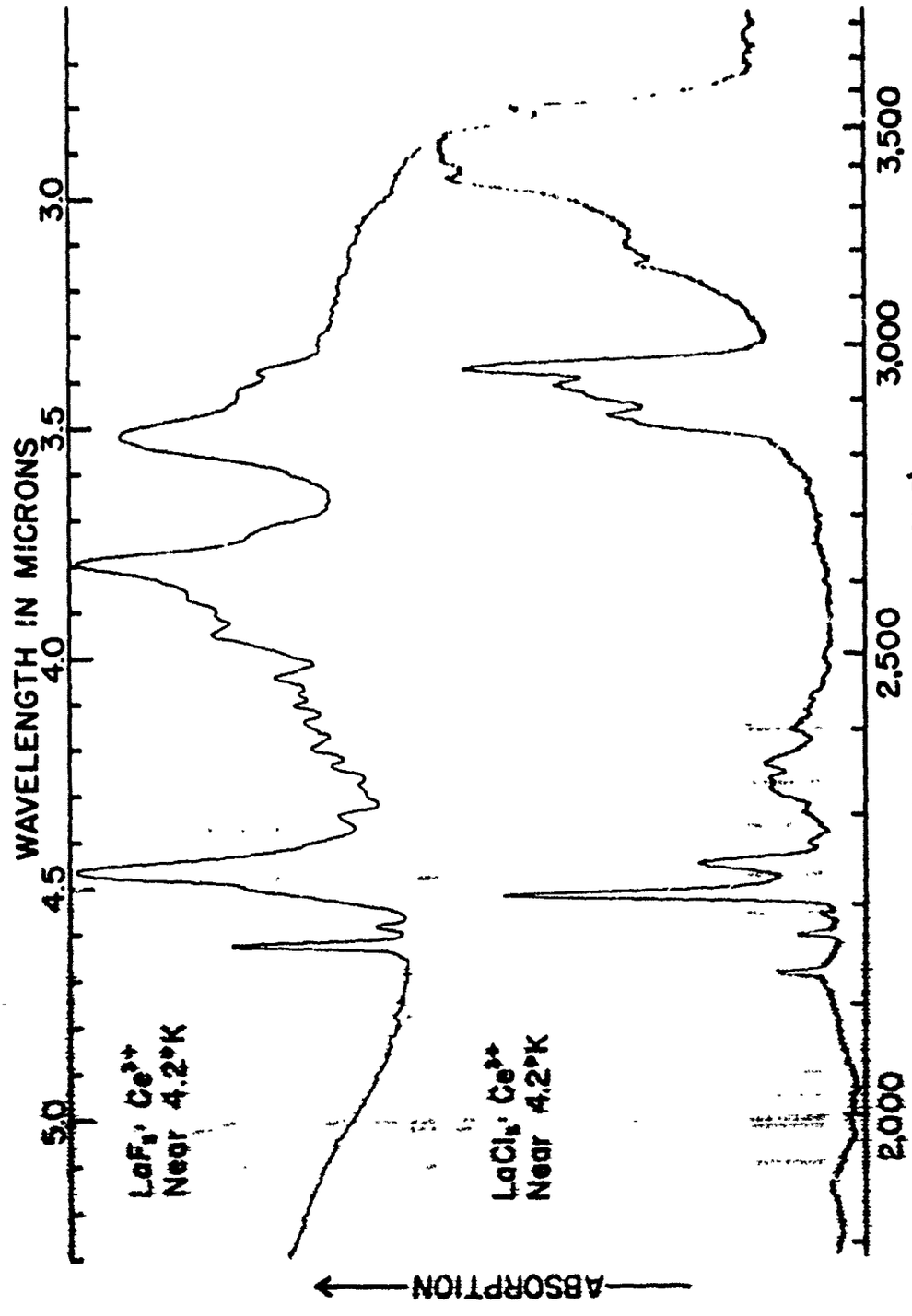


Fig. 4 The spectra of Ce^{3+} in LaF_3 and $LaCl_3$ near 4.2°K

It was mentioned previously that the weak broad lines between about 2200 and 2500 cm^{-1} were due to interactions of lattice frequencies with the electronic transitions. Figure 5 contains a list of the lattice frequencies which have been observed experimentally in LaF_3. Three possible electronic transitions have been chosen at 2160, 2180 and 2235 cm^{-1}. Each of the lattice frequencies has been added to the electronic transition and compared with the observed spectrum. Good agreement between the calculated and observed vibronic lines is noted. Similar results are obtained for the $LaCl_3$ spectrum [12, 13].

Although Ce^{3+} in $LaCl_3$ and LaF_3 has a simple electronic configuration, its spectrum in these crystals is still far from being understood. Many interesting problems involving the crystal field effects on rare earth ions remain to be solved. Our group at the Naval Ordnance Laboratory at Corona is continuing its endeavors toward the solution of some of these problems.

Lattice Vibrations (cm^{-1})		Electronic Transitions (cm^{-1})			
			2160	2180	2235
176	IR	Cal.	2336	2356	2411
		Obs.	2335	2357	2412
210	IR		2370	2390	2445
			----	2390	2447
277	IR		2437	2457	2512
			2434	2459	2515
293	R		2453	2473	2528
			2447	2469	2528
358	IR		2518	2538	2593
			2515	2535	----
368	R		2528	2548	2603
			2528	----	2604

Fig. 5 Table of vibronic lines for Ce^{3+} in LaF_3

References

(1) G. Burns, Phys. Rev. 128, 2121, (1962)

(2) C. K. Asawa and R. A. Satten, Phys. Rev. 127, 1542, (1962)

(3) H. S. Jarret, J. Chem. Phys. 31, 1579, (1959)

(4) R. J. Lang, Can. J Research 14, 127, (1936)

(5) E. Y. Wong and I. Richman, J. Chem. Phys. 36, 1889, (1962)

(6) I. Richman and E. Y. Wong, J. Chem. Phys. 37, 2270, (1962)

(7) C. A. Hutchison, Jr. and E. Y. Wong, J. Chem. Phys. 29, 754 (1958)

(8) J. M. Baker and R. S. Rubins, Proc. Phys. Soc. 78, 1353, (1961)

(9) J. Murphy, H. H. Caspers and R. A. Buchanan, Navweps Report 7237, Quarterly Report: Foundational Research Projects, July-Sept. 1962, Page 111

(10) G. H. Dieke, "Spectroscopic Observations on Maser Materials" in Advances in Quantum Electronics, J. R. Singer, ed. (Columbia Univ. Press (1961))

(11) Z. J. Kiss, Phys. Rev. 127, 718 (1962)

(12) R. A. Buchanan, J. Murphy and H. H. Caspers, Navweps Report 8141, Quarterly Report: Foundational Research Projects, Oct-Dec. (1962)

(13) R. A. Buchanan, J. Murphy and H. H. Caspers, Bull. of Am. Phys. Soc. 1962 Winter Meeting in the West, R10, p. 616 (1962)

SOLID STATE MICROWAVE AMPLIFIERS

Edward T. Hooper, Jr.
U. S. Naval Ordnance Laboratory
Silver Spring, Maryland

Introduction: The Problem

The new and expanding technology of solid state microwave amplifiers holds promise of improving the range and sensitivity of search and fire-control radars and of high frequency communications systems. The superhetrodyne receiver technique is still the most commonly employed method in all of these fields. The incoming signal from the antenna goes through a preamplifier, following which it is mixed with a variable frequency local oscillator signal to produce a fixed intermediate frequency signal which can be amplified efficiently. The preamplifier is highly desirable for sensitivity, selectivity, and the prevention of unwanted radiation of the local oscillator signal. Most microwave systems have no such preamplifier. The few which do generally employ a traveling-wave-tube amplifier to perform this task. In order to amplify the signal before encountering the noise of the mixer stage it is desirable, in fact necessary, to have low internal noise in the preamplifier stage.

Several solid state devices have been proposed as solutions to this problem. Without exception, however, they turn out to be complex, or unorthodox, or downright exotic. In any event their inherent promise has been slow in fulfillment chiefly due to the lack of basic knowledge of these "new" amplification processes and lack of engineering solutions to the new problems they present. To better understand some of the difficulties involved, first consider the amplification process itself.

Amplification

An amplifier is "a device for increasing the power associated with a phenomenon without appreciably altering its quality, through control by the amplifier input of a larger amount

of power supplied by a local source to the amplifier output." [1]

The local source of power imparts energy of motion to some "agent" such as the free electron in a vacuum tube. Then by geometry or circuitry a little energy affects the movement of the agent and larger amounts of energy are released to an external circuit. Herein resides the nonlinearity necessary for amplification.

While the local sources of power fall into two classes, direct power and alternating power, the agents employed are many and varied. The proper choice of power source depends on the agent employed.

In Figure 1 amplifiers are classified as to their local power and agent employed. In the first group the agent is free, that is, has translational energy of motion, and as a consequence is excited by direct local power. In the second group the agent is bound or constrained and is excited by alternating local power.

Solid state amplifiers such as the Transistor, Tunnel Diode and Technetron, are included in the first group. Of these, only the Transistor and Tunnel Diode are presently just beginning to operate in the low microwave frequency range. Our interest then lies chiefly with the amplifiers using alternating local power. Of these the solid state parametric diode and parametric magnetic amplifiers are the principle contenders for a solution of our microwave amplifier problem, the solid state maser being somewhat too elegant for this application. Although the principles discussed apply to both, particular emphasis will be on the magnetic type of parametric amplifier.

When the local power is direct only one frequency, the signal frequency, is present. If the input-output characteristic is linear, no new frequencies are developed. With alternating local power the situation is fundamentally different. Two frequencies, the signal and the local power frequencies, are now present. In the presence of the nonlinear device modulation takes place, cross product terms appear in the equations for current or voltage, and new frequencies appear. This is a necessary but not sufficient condition for parametric amplification. It is not sufficient, obviously, since there are modulators with power gains less than unity. Figure 2 illustrates the set of new frequencies produced by the amplitude modulation process, the familiar upper and lower sidebands about the local power frequency or pump frequency, ω_p. The presence of these and higher order modulation products require that special consideration be given to the parametric amplifier. There are multiple output ports in these amplifiers and under certain conditions multiple input ports. Frequency translation is inherent in the amplification process, i.e., the output power may be at a different frequency than the

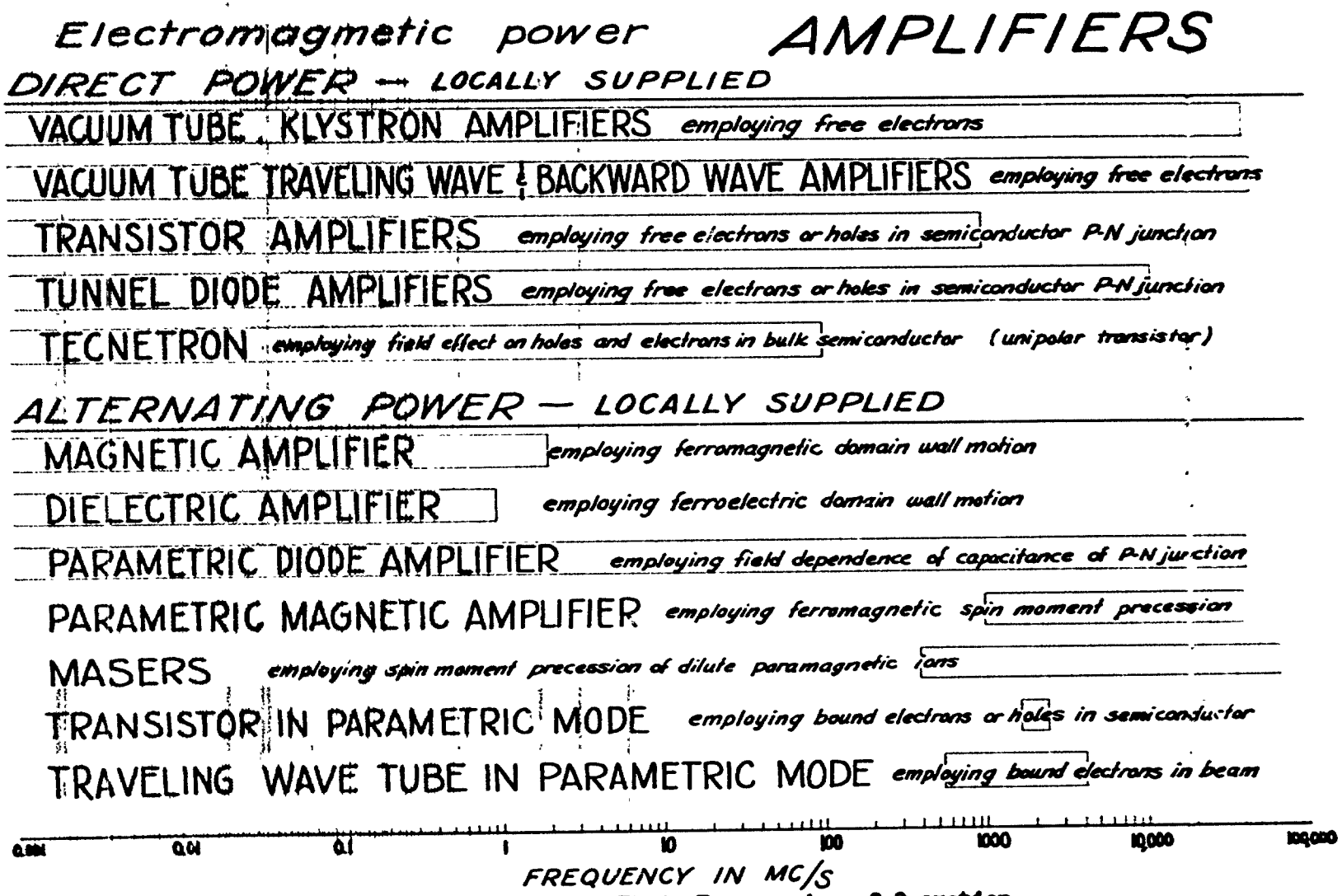


Fig. 1 Classification of Amplifiers and Their Frequencies of Operation

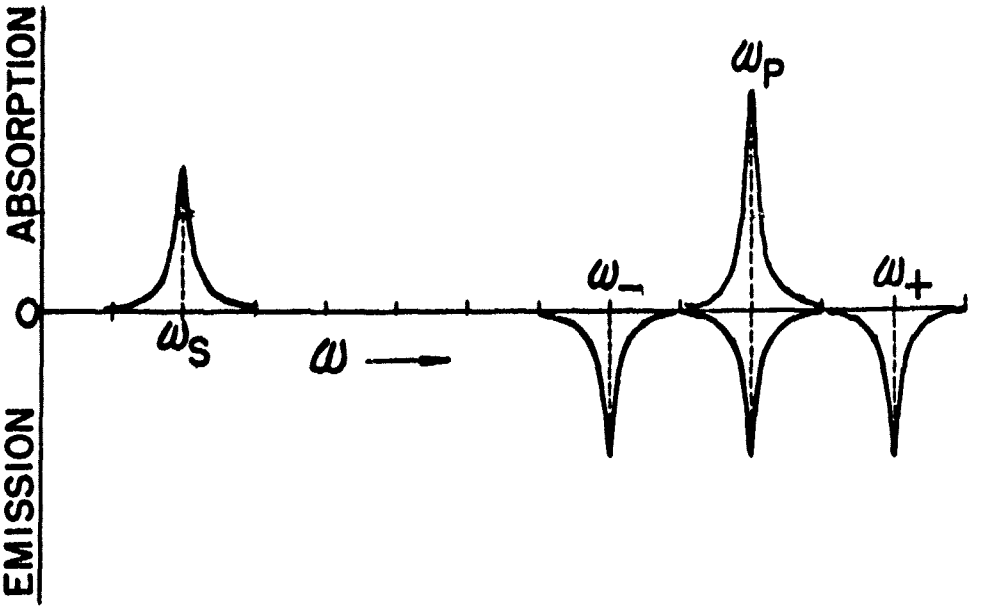


Fig. 2 Absorption and emission in the amplitude modulation process

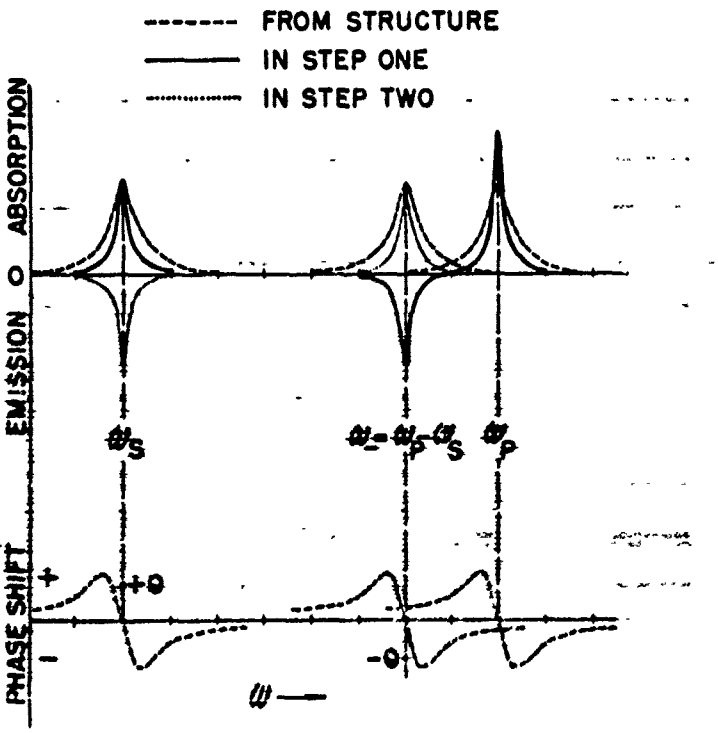


Fig. 3 Feedback process in the lower sideband case

input power and there are constraints on the conversion of power from one frequency to another.

Parametric Phenomena

The relationships between powers at different frequencies were first derived for the nonlinear reactance by Manley and Rowe.[2] These sets of equations are conservation of energy equations written as sums of power over frequency ratios. The essential properties of nonlinear (actually time varying) reactances which give rise to parametric phenomena are that the reactance varies with voltage, the impedance varies with frequency and there is a sign change in the power over frequency ratio associated with the lower sideband which signifies that the modulator delivers power to the source. The losslessness permits low noise operation.

Figure 3 illustrates modulator operation when the upper sideband power is zero. Regeneration occurs in two steps by generation of the lower sideband and the action of the lower sideband and pump frequencies in producing a new lower sideband at the signal frequency. The phase shift is correct for reinforcing the original signal. When the lower sideband power is zero a similar argument yields an additional absorption rather than emission at the signal frequency. Thus no regeneration takes place and the power gain is a stable finite value equal to the ratio of upper sideband to signal frequencies. If power is present in both sidebands other interesting effects are obtained.[3]

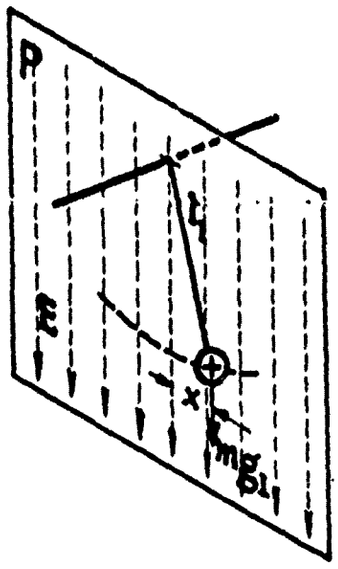
The physical mechanisms behind these parametric phenomena can be made more tangible by considering a resonant system such as the pendulum of Figure 4. The pendulum with a positively charged bob is oscillating in the plane P. Also in this plane is the electric field E. By varying the electric field E the motion of the pendulum can be varied. Without the electric field the equation of motion for small amplitudes is

$$\frac{d^2x}{dt^2} + \frac{g_1}{l_1} x = 0$$

The periodic modulation of the electric field is equivalent to a periodic modulation of the gravity coefficient g_1 and therefore constitutes the "parametric" term in this "parametric" equation. The angular frequency in radians per second is ω_1. The pendulum could also be pumped by varying l_1. This is the method used by a monkey swinging from a tree limb or a child in a swing when he wishes to increase the amplitude of swing. If

$$g_1 = g_0 + g_p \cos \omega_p t$$

where g_0 = normal gravity field coefficient unperturbed



$$\omega_1 = \sqrt{\frac{g_1}{l_1}}$$

Fig. 4 Parametrically pumped pendulum

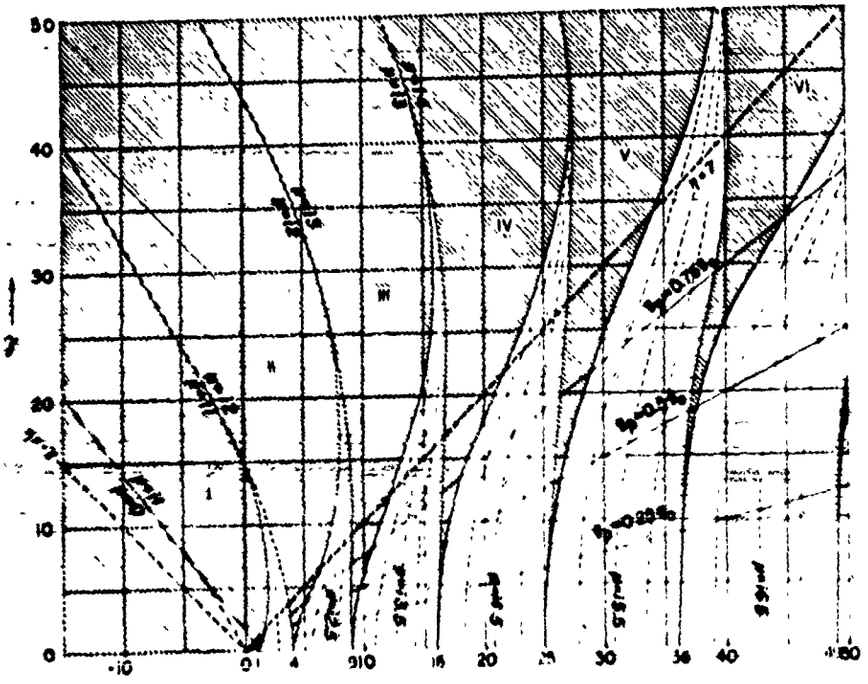


Fig. 5 Stability chart of Mathieu's equation

$g_p \cos \omega_p t$ = apparent perturbation of the gravity field coefficient due to electric field.

Then

$$\frac{d^2x}{dt^2} + \frac{g_o + g_p \cos \omega_p t}{l} x = 0$$

This is a form of Mathieu's equation.[4] Figure 5 is a plot of the stable and unstable solutions of this equation where

$$\eta = \frac{4g_o}{l\omega_p^2} \qquad \gamma = \frac{4g_p}{l\omega_p^2}$$

The complex energy propagation constant $\mu = \alpha + i\beta$ where α is attenuation constant, β is phase constant. In the blank regions μ is imaginary, there occurs no attenuation of energy and any oscillation in the pendulum is stable. These are pass bands. In the shaded regions μ is complex or real, attenuation results and energy is absorbed by the pendulum. In the absence of losses the pendulum oscillations increase without bound and an unstable pendulum results. These are stop bands. They are labled I, II, III, IV, etc. Operation is confined to the region to the right of and below the line $\eta = \gamma$. For $\gamma = 0$ there is no periodic perturbation. As the perturbation term increases (g_p increases) operation occurs further up in the map. Note that stable oscillations are the norm and pumping occurs at only a few places.

An intriguing device is the inverted (or unstable) pendulum. This device operates above the $\eta = \gamma$ line where instability is the norm but where stable operation can be obtained if the pendulum is pumped at the proper frequency.

The pendulum of Figure 4 can be pumped in the instability regions I, II, III, IV, etc. The pump and system resonant frequency relations are given in Table I for these instability regions.

TABLE I

PUMP AND SYSTEM RESONANT FREQUENCY RELATIONS

Instability Region	η (for $\gamma = 0$)	ω_p, ω_o Relations (for $\gamma = 0$)	ω_p, ω_o Relations (for $\gamma = \eta$)
I	1	$\omega_p = 2\omega_o$	$\omega_p = 1.8\ \omega_o$
II	4	$\omega_p = \omega_o$	$\omega_p = 0.89\ \omega_o$
III	9	$\omega_p = \frac{2}{3}\ \omega_o$	$\omega_p = 0.62\ \omega_o$
IV	16	$\omega_p = \frac{1}{2}\ \omega_o$	$\omega_p = 0.46\ \omega_o$
V	25	$\omega_p = \frac{2}{5}\ \omega_o$	$\omega_p = 0.36\ \omega_o$

Pumping in the first instability region is the usual mode employed by a person in a swing who desires to pump the swing to greater amplitudes. He raises and lowers his body center of mass and thereby changes the "pendulum" length at a frequency double the swing frequency. Pumping in the other instability regions is also possible as is pumping using harmonics of the pump frequency. Figure 5 indicates these permitted frequencies for pumping a system resonant at ω_o.

Since $\omega_1^2 = \frac{g_1}{l_1}$ and $g_1 = g_o + g_p \cos \omega_p t$, the frequency ω_1 is now complicated and for operation in region I consists of the unperturbed resonant frequency ω_o and two new frequencies, $\omega_o + \omega_p$ and $\omega_o - \omega_p$ as shown in Figure 7. The (1) denotes the pendulum under discussion. These frequencies will be utilized once coupling effects are discussed.

If a second pendulum is added, as in Figure 8, also parametrically pumped, coupling can be assumed to occur between them. Physically the coupling could be through a number of mechanisms, e.g., torsion of the common support of the two pendulums. This is usually described by normal mode theory. A secular equation is written, solved, and the two roots are the two normal mode frequencies of the system. There is a slight shifting of the normal mode frequencies due to the coupling. If the two pendulums both have the same resonant frequency, or are "tuned" to each other, their amplitudes of swing follow a pattern as shown in Figure 9. Energy of oscillation is removed from pendulum 1 at one point and is all stored in pendulum 2. Later this condition is reversed. If the pendulums are detuned

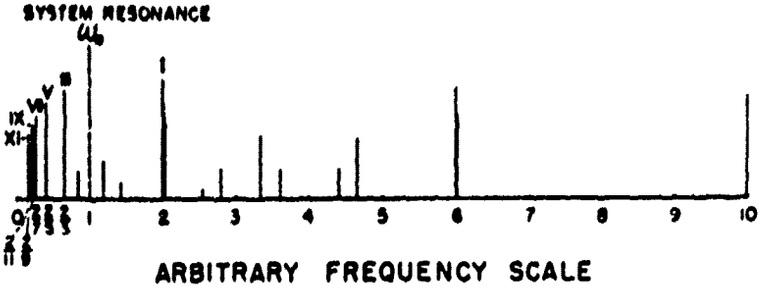


Fig. 6 Permitted pumping frequencies

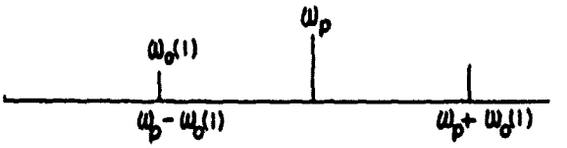


Fig. 7 Pumping in first instability region at $\gamma = 0$

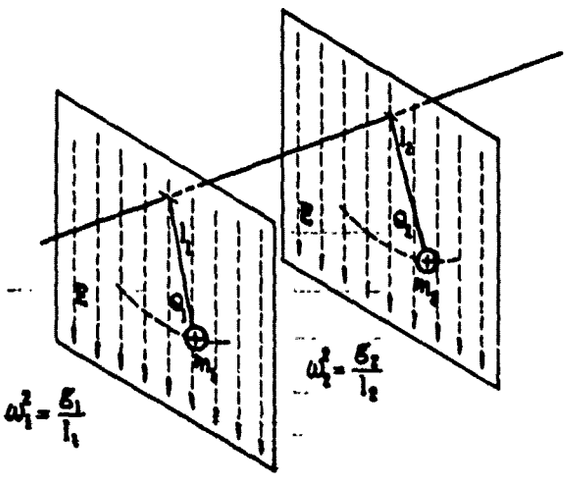


Fig. 8 Two parametrically pumped-position coupled pendulums

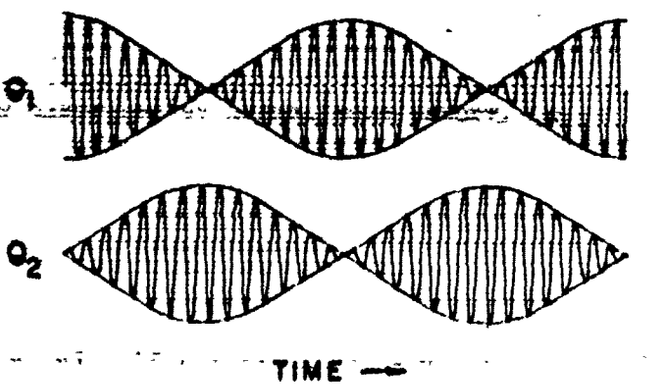


Fig. 9 Plot of positions of two tuned coupled pendulums

performance of Figure 10 is obtained where an incomplete transfer of energy is indicated. Coupling is reduced when the pendulums are detuned.

The ω_1 and ω_2 are not single frequencies for $g_p>0$ and for pumping in instability region 1 both consists of an array of frequencies as was shown in Figure 7. If pendulum 1 is detuned from pendulum 2 so that $\omega_0(1)$ lies slightly below $\frac{\omega_p}{2}$ while $\omega_0(2)$ lies an equal distance above $\frac{\omega_p}{2}$, the spectrum of Figure 11 results. For slight detuning both $\omega_0(1)$ and $\omega_0(2)$ can lie in the pumping band.

Conditions are now satisfied for parametric amplification to take place in the system. An oscillation at $\omega_0(1)$ produces $\omega_p - \omega_0(1)$ which couples strongly to $\omega_0(2)$. This $\omega_0(2)$ produces $\omega_p - \omega_0(2)$ which in turn couples strongly back to $\omega_0(1)$ reinforcing the original oscillation. For a strong pump, oscillations occur at these frequencies. With reduced pump power an oscillation at $\omega_0(1)$ would be amplified. A signal $\omega_0(2)$ would also be amplified as this type of parametric amplifier is a so-called double channel amplifier. These effects are discussed at greater length elsewhere.

Parametric Magnetic Amplifier

The pendulum is not the only example of a resonant system. There are many others which in principle could be employed to construct a parametric amplifier. One is the resonance of a magnetic spin system.

In Figure 12 the figure on the left shows the magnetic moment $\vec{m}$ of an electron. The vector $\vec{L}$ is the angular momentum of the electron in orbit, $\vec{S}$ is the angular momentum of the electron spin and $\vec{J}$ is the total resultant angular momentum. The magnetic moments $\vec{m}_L$ and $\vec{m}_S$ result from the orbital and spin angular momentum. The $\vec{m}$ is the resultant magnetic moment. In the presence of a constant magnetic field $\vec{H}_0$ the magnetic moment vector, $\vec{m}$, precesses about $\vec{H}_0$ as shown. Losses which damp this precession cause $\vec{m}$ to spiral into the direction of $\vec{H}_0$. The undamped precessional angular frequency is ω_0. If an alternating magnetic field of frequency ω_0 is added in the x-y plane, the alternating field will pump the precession and reduce or overcome the losses changing or maintaining a constant angle between $\vec{m}$ and $\vec{H}_0$. If the alternating field frequency differs from ω_0 energy is not absorbed by the system. When it equals ω_0 strong absorption occurs. The resonance phenomena obtained is equivalent to pumping in the instability region II and the alternating field frequencies generally fall in the microwave frequency region.

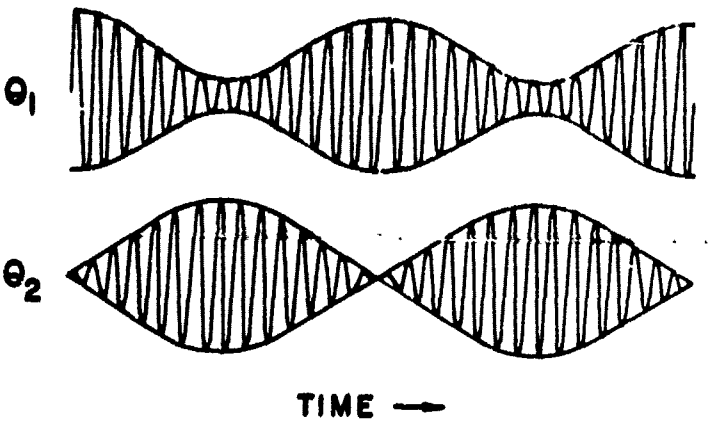


Fig. 10 Plot of positions of two slightly detuned coupled pendulums

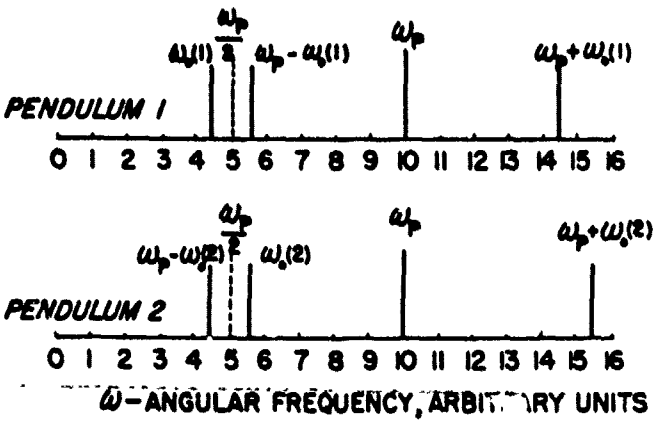


Fig. 11 Spectrum of coupled pendula feedback process

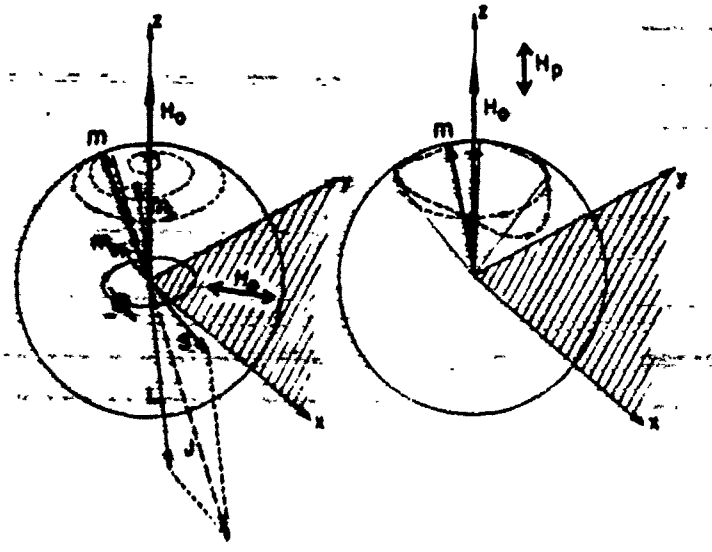


Fig. 12 Perpendicular-parallel pumping of ferromagnetic spin system

It has recently been found that pumping is also possible when the alternating field is parallel to the constant field H_0 as shown in Figure 12, right side. Since $\vec{M}$ has constant length and the end traces its path on the surface of an imaginary sphere as shown, any anisotropy or difference in characteristics between the x and y directions results in an elliptical precession path the extreme ends of which turn down along the surface of the sphere and result in a change in the Z component of $\vec{M}$. This component can be pumped by the field H_p and characteristically is pumped in instability region I, i.e., where $\omega_p = 2\omega_0$.

Thus far we have only one resonance and more are needed to construct our magnetic parametric amplifier. The magnetic resonance phenomena is more complicated however than so far indicated. If all of the individual spins have the same precession angle and are in phase the uniform precessional mode results as in Figure 13(a). If there is a difference in phase in the precession of the moment vectors, waves appear to move through the medium as in Figure 13(c) traveling in the direction of the constant magnetic field, perpendicular to it or at some arbitrary angle. These excited states are spin waves. If the spin wave length is sufficiently short the waves can be represented as plane waves with propagation constant k, in an infinite medium with negligible effect from the sample boundary. As $k \to 0$ and the wave length λ approaches infinity the assumptions leading to a plane wave equation are violated. Principally the boundary conditions now become important; a plane wave representation is not correct, the exchange forces exert only a slight influence on the motion of the spin; the characteristic frequency is then nearly independent of the exchange field and consequently is nearly independent of k. The frequency is now dependent on the long range dipolar interaction of the spins and the static magnetic field. As wave lengths approach sample size the correct normal modes are the magnetostatic modes or Walker modes. These "standing waves" depend on the shape of the body but not its size, provided the body is large enough for exchange torque to be neglected yet small enough for propagation effects to be neglected. These standing waves are illustrated in Figure 13(b). Figure 14 shows the region of spin wave excitation (plane wave for $k > k'$) for spin wave propagation angle θ_k with respect to the static field H_0. The shaded area is the approximate representation of the density of magnetostatic modes for a sphere.

A photograph of an array of the magnetostatic modes of a single crystal sphere of yttrium iron garnet produced when the magnetic field is varied is shown in Figure 15. The large absorption (110) is the uniform precessional mode. The identifying numbers are Walker's nomenclature.[6] These resonances can be used to satisfy the necessary conditions for parametric amplification.

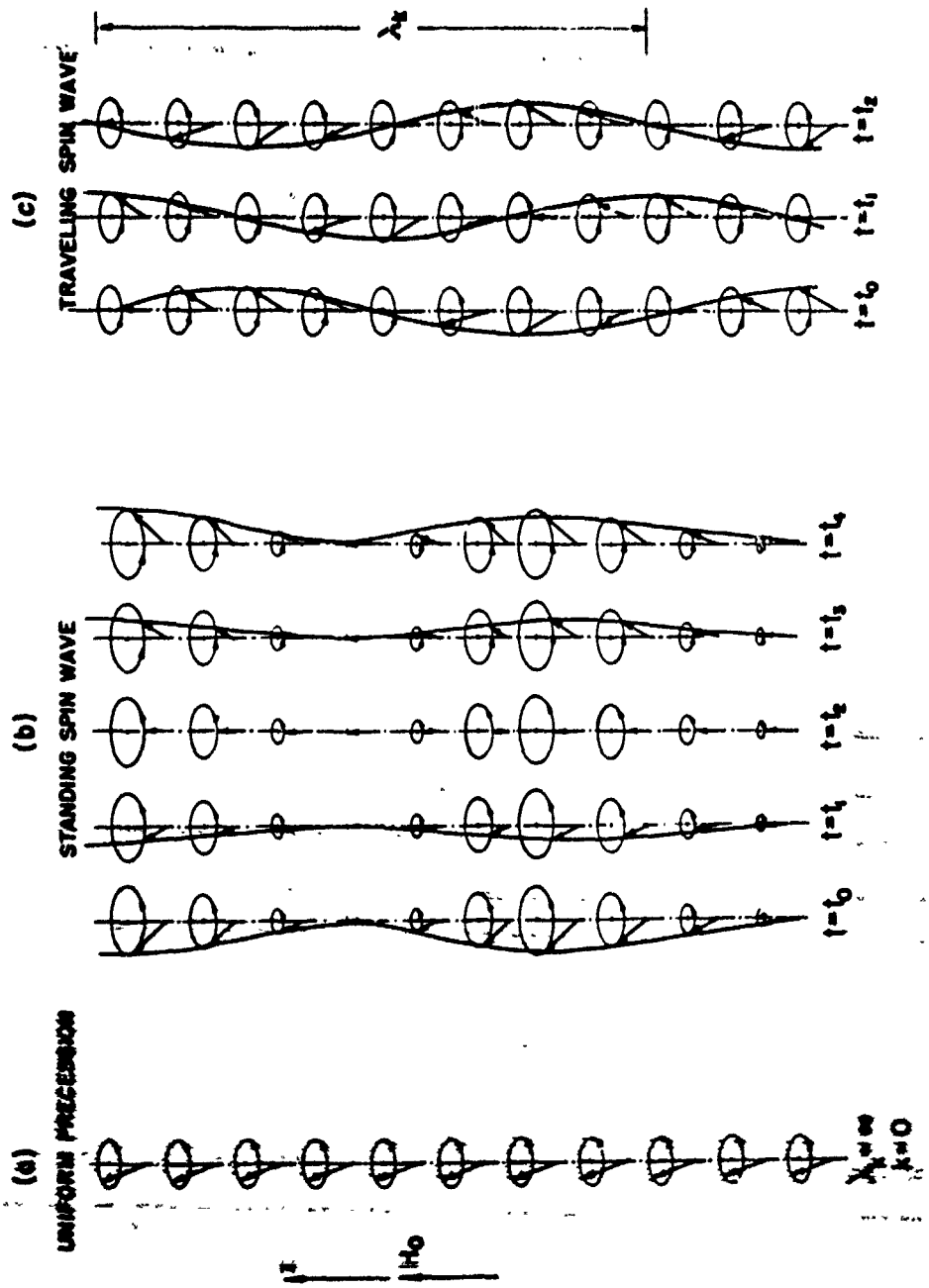


Fig. 13 Spin Waves

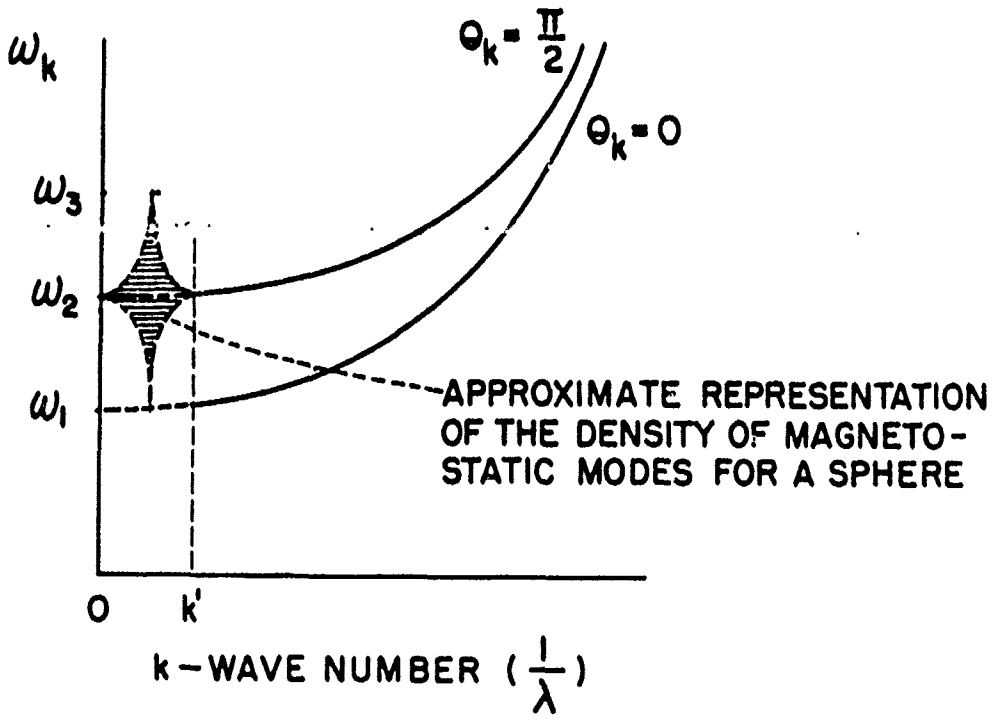


Fig. 14 Spin wave manifold

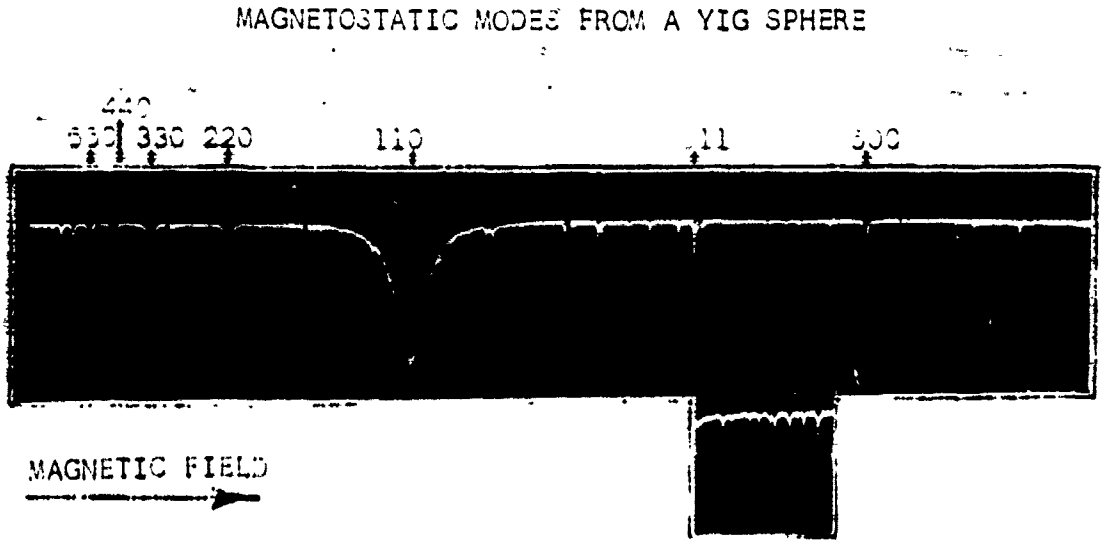


Fig. 15 Magnetostatic modes

Amplifiers have been developed by Suhl[7] using perpendicular pumping and Denton[8] using parallel pumping. The parametric conditions can be satisfied in one of several ways. The system can provide cavity resonances at signal and idler frequencies and employ the uniform precession. This has been termed the electromagnetic mode by Suhl. Similarily one cavity and two magnetostatic modes can be employed. This is called the semistatic mode. Or, three magnetostatic modes can be employed resulting in the "magnetostatic" mode of operation. Denton employed two magnetostatic modes whose frequencies sum to the pump frequency. They were closely spaced about a frequency equal to half the pump frequency. This is pumping in instability region I.

In the spin wave manifold of Figure 14 certain magnetostatic modes which are being pumped are degenerate in frequency with an array of spin waves with high k numbers. Consequently, pumping energy is lost to the spin wave system and potentially noisy spin wave systems are excited.

By pumping in other instability regions the pump frequency can lie outside of the spin wave manifold. The desired magnetostatic modes can be excited while undesired spin waves are not excited. A potential reduction in noise is then possible. Pumping in instability region III has been demonstrated where the pump frequency is two-thirds of the average of the two system magnetostatic modes.

Spin waves are also excited by small imperfections in the surface of the single crystal yttrium iron garnet. These broaden the resonant line widths. Careful grinding and polishing is necessary to reduce the line widths to the intrinsic line width of the material. Figure 16 is a photograph of a polished sphere held on the end of a magnet. The head of a straight pin at the top gives an indication of the size. The rough piece is an unpolished YIG crystal from which a sphere is ground.

In the ferromagnetic microwave amplifier under development at the Naval Ordnance Laboratory this crystal is mounted in a quartz tube spaced off the end wall of a rectangular microwave cavity operating in the TE_{011} mode. The cavity and associated equipment is shown in Figure 17. Pump power enters through the iris window in the cavity and establishes a microwave magnetic field H_p. Parallel to this field is the static magnetic field H_o. The signal input enters through the coaxial line and single turn of wire wrapped about the quartz tube and sphere. The device operating as a negative resistance amplifier produces more power at the signal frequency than the input power. This power passes down the coaxial line and through the action of the circulator appears at the output port.



Fig. 16 Polished single crystal sphere

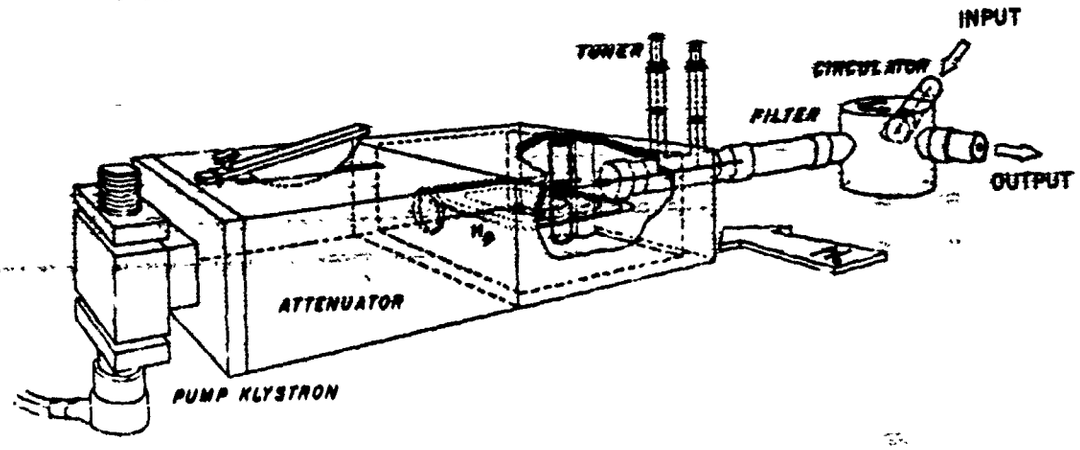


Fig. 17 Ferromagnetic microwave amplifier

The pumping is parallel. Two magnetostatic modes are employed and pumping in the instability region III is utilized to reduce noise.

Conclusions

Some of the novel features of solid state parametric amplifiers have been discussed. These must be properly understood and utilized in designing these devices into existing systems. The general theory of parametric pumping is applied in the case of the ferromagnetic microwave amplifier in an effort to reduce the noise encountered in this amplifier.

References

1. American Standards Association definition

2. J. M. Manley and H. E. Rowe, "Some General Properties of Non-Linear Elements- Part I - General Energy Relations" Proc. IRE Vol. 44, No. 2, pp 904-913, July 1956

3. D. K. Adams, "An Analysis of Four Frequency Non-Linear Reactance Circuits" Trans. on Microwave Theory and Techniques, Vol MTT-8, No. 3, May 1960

4. Leon Brillouin, Wave Propagation in Periodic Structures (book) Dover Publications 1946

5. E. T. Hooper, "The Nature of Microwave Parametric Amplification" U. S. Naval Ordnance Laboratory Technical Report 62-102, 5 Sept 1962

6. L. R. Walker, Phys Rev 105, 390-399 (1957)

7. H. Suhl, JAP 28 1225-1236 (1957)

8. R. T. Denton, JAP 32 3005 (1961)

PROPERTIES AND DEVICE APPLICATIONS OF EVAPORATED InSb FILMS

H. H. Wieder
U. S. Naval Ordnance Laboratory
Corona, California

ABSTRACT

The galvanomagnetic properties of evaporated intermetallic semiconductor films of indium antimonide were investigated and are presented with emphasis on their application to devices employing the Hall effect. The design and construction of a magnetometer based on the pulsed operation of a thin film Hall generator will be described in detail. The magnetometer is capable of resolving d.c. or alternating magnetic fields of the order of 10^{-4} gauss. Operation of either d.c. or pulse driven analog multipliers, dividers, square root computers and function generators based on the Hall effect in InSb films will be discussed and compared to similar devices using bulk crystalline materials.

INTRODUCTION

Vacuum deposited thin films of indium antimonide with galvanomagnetic properties similar to those of bulk crystalline InSb can be prepared most advantageously[1-3] by means of the "three temperature control process."

The main feature of this process is the individual regulation and temperature control of separate crucibles containing purified and zone refined In and Sb. Microscope cover slips of glass used as substrates for the film are heated by a radiant source under close temperature control. The temperature of the substrate must be maintained above the condensation temperature of either the In or the Sb within the vacuum chamber and below the melting temperature of InSb. The rate of evaporation in conjunction with the effective substrate temperature determines to a large extent the stoichiometry of the deposited film and the size of the crystallites within it.

The impurities incorporated into the film during deposition

determine the type, concentration and the effective mobility of the charge carriers. Although the films are polycrystalline and there is evidence of localized variations in their composition and a fluctuation in their thickness, effective electron mobilities of the order of 10^4 $cm^2/(volt\text{-}sec)$ have been obtained in 10^{-4} cm thick films.

The high mobility and the long mean free path of electrons in InSb suggest a large number of potential applications for these films, especially with continued improvement in the methods of crystallization and deposition. At present, the direct application to devices based upon the Hall effect appears most promising. However, the galvanomagnetic properties of these films need to be known in order to build practical Hall generators. In the following sections details will be presented on some measured galvanomagnetic properties of films. A comparison will then be made with the known electrical properties of bulk crystalline InSb and finally, the development of some specific circuitry for selected applications of thin film Hall generator applications will be illustrated.

EXPERIMENTAL

If the rate of evaporation of each of the film constituents is held constant it is then reasonable to expect the average thickness to be dependent upon the duration of the evaporation cycle. Figure 1 shows that this is indeed the case for evaporation periods up to 20 minutes. Films up to 5×10^3 Å thick were measured by means of the interferometric technique described by Tolansky.[4] For thicker films the infrared interference reflection measurements described by Potter and Kretschmar[2] were used. Both methods yield at best a nominally average value of thickness, but each appears to be self-consistent within an experimental error estimated to be of the order of 4%. Films less than 10^{-4} cm in thickness contain generally a large number of pinholes and sequentially evaporated films, deposited under apparently identical evaporation procedures, vary in thickness by more than 100%. Thicker films are pinhole free, appear to be better crystallized than the thin films and their x-ray diffraction pattern suggests complete stoichiometry of the compound. Films in the thickness range of 10^{-4} to 2×10^{-4} cm appear to have more reproducible electrical characteristics and the subsequently described experiments apply primarily to such films. Following deposition of InSb, the coated glass slides are cut to either of the shaped contours shown in Fig. 2. Gold electrodes are then evaporated for the Hall and drive current electrodes and leads are attached to these by means of silver adhesive paste.

The electrostatic shorting effects of the electrodes were investigated for both geometries shown in Fig. 2. A decrease in the Hall voltage of about 10% was noted for wide area drive current electrodes compared to point contact electrodes.

By decreasing the length of a Hall generator i.e., extend-

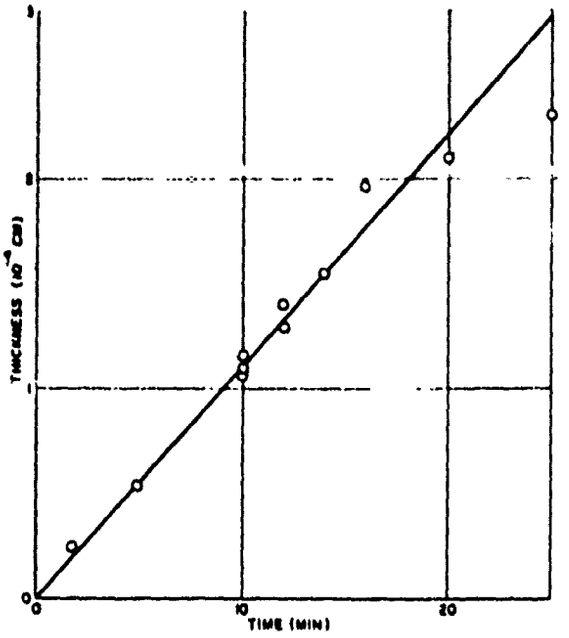


Figure 1. Film thickness is shown to depend upon the duration of evaporation for one series of sequentially deposited films.



Figure 2. Two different InSb film Hall plates on microscope cover glass substrates are shown with evaporated gold electrodes. Pilot holes are for wire leads.

ing the drive current electrodes while maintaining the width constant, the theoretical predictions[5] of electrostatic shorting were tested for three different specimens. Figure 3 shows the excellent adherence between the theoretical curve and the experimental data. The subsequently described measurements were performed on Hall generators having a length to width ratio of 2.8.

GALVANOMAGNETIC MEASUREMENTS

Figure 4 shows the Hall voltage v_h as a function of drive current i of two different Hall generators subjected to an identical magnetic field of 5×10^3 gauss. Sample 1b is typical of the undoped InSb films fabricated by means of the procedures described above; v_h is linear in i up to 6 ma while higher currents cause significant joule heating of the film. The increase in temperature affects the Hall coefficient R_h and the conductivity σ primarily through the thermal activation of charge carriers. For n-type films R_h decreases with temperature hence the nonlinear portion of v_h vs. i in Fig. 4 shows a slope less than one. On the other hand, sample 88a is an InSb film deliberately doped with copper during the evaporation process. Its Hall voltage is linear in i up to 9 ma but thereafter it increases sharply with the drive current. The differences in the behavior of the two samples may be understood in terms of the experimentally measured behavior of R_h as a function of temperature. Figure 5 shows a plot of R_h vs. the reciprocal absolute temperature, with R_h calculated from the linear portion of graphs such as shown in Fig. 4. Sample 1b exhibits a behavior typical of n-type InSb[6] with donor activated electronic conduction in the extrinsic temperature region. Sample 88a is shown to behave in a manner similar to bulk p-doped crystalline InSb. In the extrinsic temperature region its charge carriers are predominantly acceptor activated holes with copper probably replacing In substitutionally in the InSb crystal lattice and acting as a doubly ionized acceptor. At 300°K, it is clear therefore from Fig. 5 that the slope of R_h with temperature is positive for sample 88a while it is negative for sample 1b.

In general, all films of InSb evaporated by means of the "three temperature control process" are n-type if acceptors are not deliberately introduced into the heated crucibles or the vacuum chamber. The sequentially evaporated films do, however, show a considerable variation in the degree of charge carrier compensation and in the number of donor impurities present, irrespective of the stoichiometry. Since for n-type InSb, $R_h \sim (1/ne)$, deliberate doping of a film with donor impurities reduces the temperature dependence of R_h.

The Hall coefficient of an InSb film having both positive and negative charge carriers may be written as:

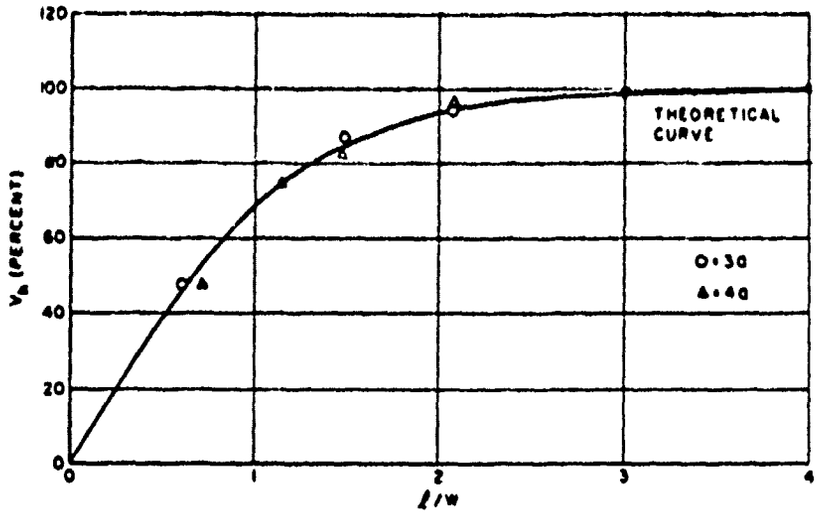


Figure 3. Percent of Change in Hall Potential as a Function of the Length-to-Width Ratio for Two Samples. The curve is the theoretical curve predicted by Isenberg, Russell, and Greene.

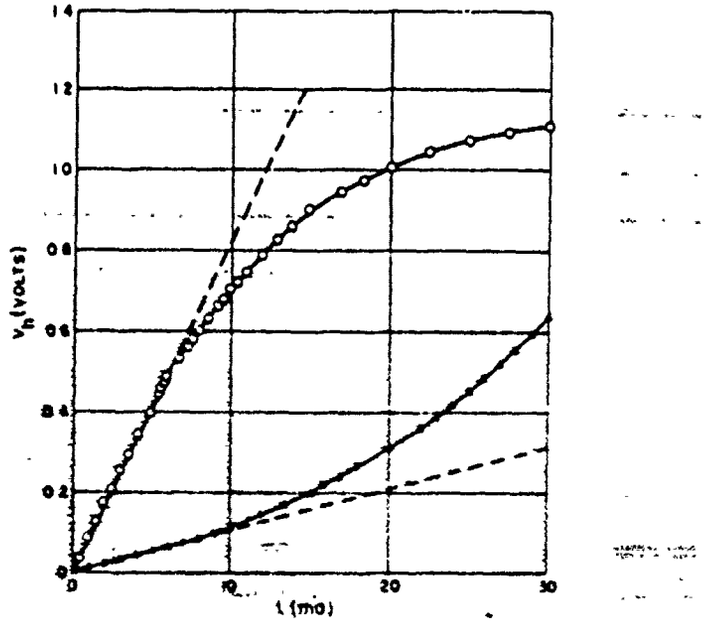


Figure 4. Hall voltage vs. Hall current for two different Hall generators in a field of 5×10^3 gauss and at an ambient temperature of 299°K; Open circles represent data obtained on sample 1b, a normal, n-type film, triangles refer to sample 88a, a copper doped p-type film.

$$R_h = (-1/a)\frac{nb^2 - p}{(nb + p)^2} \tag{1}$$

where b = the ratio of electron to hole mobilities, n = electron concentration and p = hole concentration per cm^3. Since $n = p + N$ where the total concentration of impurities is $N = N_D - N_A$ with N_D = donor concentration and N_A the acceptor concentration, it follows that:

$$p^2 + N_p - n_p = 0 \tag{2}$$

Solving for n and p yields:

$$\begin{aligned} n &= (N/2)[1 + x)^{\frac{1}{2}} + 1] \\ p &= (N/2)[1 + x)^{\frac{1}{2}} - 1] \end{aligned} \tag{3}$$

where $x = (4np/N^2)$. Substituting equation (3) in (1):

$$R_h = -(1/Ne)\frac{(\frac{1}{2})(b^2 - 1)(\sqrt{1 + x} - 1) + b^2}{[(\frac{1}{2})(b + 1)(\sqrt{1 + x} - 1) + b]^2} \tag{4}$$

In any device application of the Hall effect, it is desirable to minimize the temperature dependence of the Hall coefficient. Galavanov[7] has shown a method for reducing the temperature dependence of the Hall voltage by controlling the concentration of donor impurities in an n-type semiconductor: Suppose the Hall voltage in the extrinsic region at the temperature T_o is $v_h(T_o)$. Let the highest temperature under consideration by $v_h(T)$. Let the maximum permissible change in v_h be:

$$\frac{v_h(T_o) - v_h(T)}{v_h(T_o)} = \frac{R_h(T_o) - R_h(T)}{R_h(T_o)} = 1 - \alpha \tag{5}$$

when $\alpha = R_h(T)/R_h(T_o)$ and $R_h(T_o) = -(1/Ne)$. Since for InSb the mobility ratio $b \geq 32$ and as a first approximation may be considered to be independent of temperature:

$$\alpha = \frac{2}{(1 + x)^{\frac{1}{2}} + 1} \tag{6}$$

Calculating the minimum impurity concentration N_{min} for a predetermined temperature dependence of R_h leads therefore to:

$$N_{min} = \frac{\alpha n_i}{(1 - \alpha)} \tag{7}$$

Let the desired change in v_h remain at less than 2% between -50°C and +50°C i.e., $\alpha = 0.98$ and $N_{min} = 6.95\ n_i$; at 323°K the intrinsic carrier concentration is:

$$n_i = 6 \times 10^{14} \times 323^{3/2} \exp\left(-\frac{1510}{323}\right) \tag{8}$$

Consequently $n_i = 3.28 \times 10^{16}$ electrons/cm^3 and $N_{min} = 2.28 \times 10^{17}$ donor impurities per cm^3. The Hall coefficient in extrinsic region should not exceed a value $R_h = 27.6$ cm^3/coul. if $(\partial v_h/\partial T)_{B,i}$ is not to vary by more than 2% over the temperature range specified above. However, since the field sensitivity of a Hall generator fabricated from an n-type plate of InSb may be written as:

$$v_h/B = (1/ne)(i/d) \tag{9}$$

with B = the magnetic induction and d the thickness of the Hall plate, an increase in the total carrier concentration will bring about a decrease in its sensitivity. A compromise is thus required in device design between the desired sensitivity and thermal stability of a Hall generator.

From sample 88a in Fig. 5 the mobility ratio may be calculated[6] as b = 32 and the impurity concentration in the extrinsic region is $(N_A - N_D) = 3.91 \times 10^{17}$ impurities/cm^3. The strong temperature dependence of R_h in the vicinity of room temperature makes such doped films unsuitable for Hall generator applications. Conceivably, however, they might be useful as thermal detectors or bolometers when cooled in the vicinity of the temperature at which (dR_h/dT) is a maximum (near the transition in sign of the Hall coefficient at $R_h = 0$).

Figure 6 shows the temperature dependence of the conductivity of the above two samples calculated from the temperature dependence of their resistances measured at B = 0.

The effects of inhomogeneity in a thin semiconductor film arise primarily in the calculation of the effective resistance. Suppose that such a film consists on a microscopic scale of semiconductor grains in imperfect contact with each other. The macroscopic resistivity $\bar{\rho}$ will then appear to be larger than the resistivity ρ, of the semiconducting grains as:

$$\bar{\rho} = \rho\ [1 + \beta] \tag{10}$$

where β is a parameter that depends upon grain size and the grain geometry. The Hall constant R_h, is not materially altered by the inhomogeneity of the film. This is the case even if the semiconductor grains are surrounded by a high resistivity layer provided that this layer thickness is small with respect to the grain dimensions.

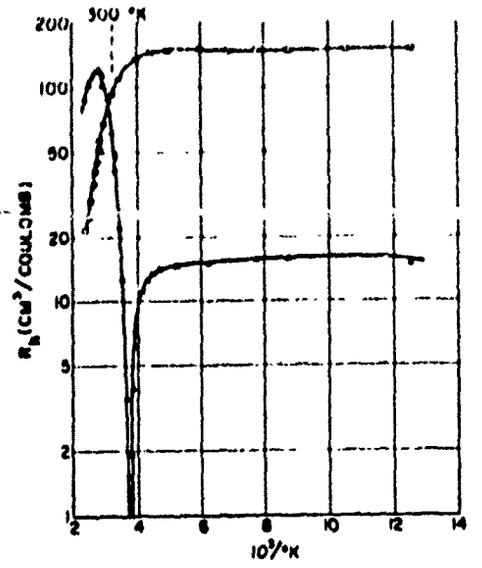


Figure 5. Hall coefficient vs. reciprocal absolute temperature for n-type sample 1b (open circles) and p-type sample 88a (triangles). Measurements were made in a constant magnetic field B = 2 x 10^3 gauss.

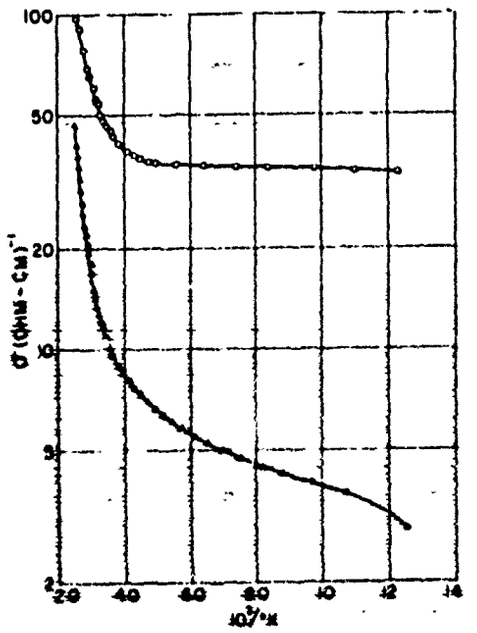


Figure 6. Temperature dependence of virtual conductivity of specimens 1b (open circles) and 88a (triangles) at B = 0.

Since the calculated thickness is at best an average value for these films and no account is taken here of intergrain barrier resistances of their temperature dependence, σ is probably best expressed as a virtual conductivity. A virtual Hall mobility μ_h may then be calculated for each of the samples by means of Figures 4 and 5.

At 300°K sample 1b has a virtual electron mobility $\mu_h = 4.95 \times 10^3$ cm^2/(volt-sec). The virtual electron and hole mobilities of sample 88a at its maximum R_h value were calculated respectively as: $\mu_n = 5.83 \times 10^3$ cm^2/(volt-sec) and $\mu_p = 1.82 \times 10^2$ cm^2/(volt-sec).

The Hall coefficient was found to be independent of the applied magnetic field over the temperature range of these investigations for sample 1b. For sample 88a the Hall coefficient decreases with magnetic field in agreement with the measurements reported on bulk InSb and the expected behavior[8] in accordance with the theoretical model of a two carrier type semiconductor. The decrease in R_h is of the order of 12% as the field is increased from 10^3 to 10^4 gauss at 300°K. A small magnetoresistive effect of the order of 20% at 10^4 gauss was measured on sample 88a with the change in resistance proportional to B^2. Other galvanomagnetic[8] effects are of second order in importance in InSb and a preliminary evaluation of their contribution to the measured adiabatic Hall potential in films in the vicinity of 300°K shows them to be negligible.

For zero magnetic flux a potential will generally appear across the output electrodes of a Hall plate when a drive current is applied to it. This potential has its origin in the misalignment of the Hall electrodes with respect to an equipotential plane across the Hall plate.

Methods of compensating for the misalignment voltage on a Hall plate generally use an external resistance bridge and the misalignment potential is reduced to zero by balancing this bridge. Another method to compensate for the misalignment voltage is to arrange a current path parallel to the Hall plate. By adjusting the resultant potentiometer circuit, the voltage across the Hall contacts at zero field can be brought down to a value less than 5μv.

Each of these methods has the advantage that very little output power will be lost in the compensating network. There are some apparent disadvantages, however. By means of an equivalent bridge circuit it may be shown that for the bridge configuration, the change in the misalignment potential with temperature is proportional to the temperature dependence of the resistivity of a Hall generator driven by a constant current source. This difficulty is not encountered in a potentiometer compensator. By carefully removing small film increments from the side arms of Hall plates such

as illustrated in Fig. 2, it was found possible to reduce the misalignment potential to a negligibly small value at room temperature. Within ±25°C of the ambient temperature the misalignment potential was found to increase linearly with the drive current within the linear v_h vs. i_h region for sample 1b. In any case the reduction of the misalignment voltage is desirable since circuital compensation methods are not very satisfactory when thermal drift of the misalignment voltage must also be suppressed.

DEVICE APPLICATIONS OF InSb FILMS

The Hall generator is a multiplier yielding the vector product of two quantities which can be translated into orthogonal electric and magnetic fields. The Hall voltage is the resultant transverse output due to the interaction of these vectors. The many applications proposed for the Hall effect are concerned, in a generic sense, with its use as a multiplier or transducer in a simple, compact, passive, solid state device with the following advantages:

(a) Hall generators depend for their operation upon majority carriers and the usual problems associated with semiconductor junction devices are not significant.

(b) Hall generators are linear devices over a considerable range of magnetic fields and input currents, a facet which is important for their use in transducers or linear detectors and demodulators.

(c) Because of the short relaxation time of carriers (of the order of 10^{-12} seconds), the Hall effect in intermetallic semiconductors such as InSb is frequency independent over the spectrum needed for probing magnetic field distribution in waveguides or cavities through the μ-wave region.

The criteria for the practical application of bulk crystalline Hall generators apply to thin film devices as well. The latter have some inherent advantages and disadvantages:

(a) Thin films mounted within the gap of a magnetic circuit such as used for Hall effect based analog multipliers, yield a large open circuit output because v_h is inversely proportional to the Hall plate thickness and also because the effective magnetic permeability and hence B increases as the gap is reduced.

(b) The large surface to volume ratio of films improves the thermal dissipation of joule heat by surface conduction and their high internal resistance (660 ohms for sample 1c) compared to bulk Hall generators (of the order of 1 ohm) simplify the design, construction and operation of associated circuits.

(c) Impurity doping may be used to decrease the temperature dependence of R_h by increasing the carrier concentration at the expense of a decrease in μ and an increase in σ for both bulk and thin film devices.

(d) The efficiency η of a film Hall generator (when matched to a load resistance equal to its internal resistance) is less than that of bulk InSb devices since η varies as μ_h^2 and the μ_h of presently available films is smaller by at least a factor of 6 compared to single crystal InSb.

(e) For a dc power dissipation fixed by the thermodynamic parameters of a Hall generator and its surroundings, a decrease in the duty cycle of i_h allows an increase in its peak amplitude while maintaining the joule heating of the film below the peak permissible thermal dissipation. Thus pulsed operation of a Hall generator offers a convenient method for increasing the effective sensitivity and extending its linear region of v_h vs. i.[9]

The inherent simplicity of the Hall effect is an obvious advantage in its use for the detection and measurement of magnetic fields. The proportionality between the magnetic induction and the Hall voltage is the basis for the magnetometer applications of the Hall effect.

The peak allowable power density P_m for a particular film geometry and conductivity is determined primarily by the joule heating. Its magnitude is the steady-state power dissipation in terms of the dc current density J_x. If the current is applied in the form of rectangular pulses of duration τ and repetition rate ν, then for an equivalent dc heating effect, $J_x = J_p(\tau\nu)^{\frac{1}{2}}$, where J_p is the peak pulse amplitude. An increase in sensitivity may thus be obtained by pulse driving a Hall detector. A further increase in sensitivity may be obtained by placing the Hall plate between ferrite or μ-metal field concentrators. The effective permeability μ_e in the gap between them is determined by the permeability and geometry of the field concentrators as well as the gap spacing.

The subsequently described results were obtained on an InSb film Hall detector mounted between ferrite field concentrators having a nominal permeability of 500. The gap between the concentrators is 0.035 cm. The dimensions of the Hall plate are: ℓ = 0.48 cm, d = 1.6×10^{-4} cm, and w = 0.24 cm. At +25°C, the conductivity of the film was determined to be $\sigma = 26.6\ (\text{ohm-cm})^{-1}$ and the Hall coefficient as $R_h = 155.3\ \text{cm}^3/\text{coul}$. The effective mobility is then $\mu = 4.13 \times 10^3\ \text{cm}^2\ (\text{volt-sec})^{-1}$. For a steady-state magnetic field, identical values of v_h are obtained either with a dc or a pulsed current drive up to a peak value of I = 9 ma. Above 9 ma, v_h still increases linearly with the pulse current J_p. The sensitivity of the magnetometer may be shown to be:[10]

$$(v_h/B) = \mu_a w\,(J_x/J_p)(P_m \mu R_h/d)^{\frac{1}{2}} \qquad (11)$$

Since for the above described film specimen $P_m = 0.336$ w/cm^2, then for a pulse duration of 10^{-7} seconds and a repetition rate $\nu = 10^3$ pps

$$v_h/B = 8.8 \times 10^{-3} \text{ volts/gauss} \qquad (12)$$

With the field concentrators in place, v_h was found to increase by a factor of 65 in the same steady state field of the order of 0.3 gauss. The effective sensitivity of the magnetometer is therefore $(v_h/B) = 0.572$ volts/gauss. Taking the smallest Hall output pulse v_h above noise as 10 μv then the smallest detectible field is 5.72×10^{-4} gauss. Experimentally this was confirmed by plotting the field strength pattern of bar magnet which yielded the expected dipole configuration. The output of the magnetometer was applied to an oscilloscope and fields out to 8×10^{-4} gauss were measured in this manner.

If the magnetic field to be measured is a periodic function of time, then a sampling method may be employed in conjunction with the pulsed Hall detector. The waveform as well as the amplitude and direction of a magnetic field may be determined and the usual advantages of sampling procedures may thus be realized. The construction and performance of such sampling magnetometers have been described earlier.[10] It was shown that the minimum detectible field is of the order of 4×10^{-3} gauss at a field frequency of the order of 1 kc. The magnetometer may be improved considerably by using thinner films of higher mobility InSb, designing the field concentrators for an optimum μ_e and bandwidth and improving the thermal heat transfer between the film and its surroundings.

Much of the literature on the applications of the Hall effect is concerned with the use of Hall generators as multipliers for analog computer applications.[11] An input parameter, translated into a dc or alternating current is applied to a Hall plate and a second input parameter is translated into an orthogonal magnetic field which is either dc or time varying. The Hall voltage output is then proportional to the product of the two input quantities and the multiplier may, of course, be used as a modulator or demodulator as well.

A prime consideration in the application of such a multiplier is the accuracy of multiplication, the range of linearity between v_h, I and B, the long term as well as the short term stability of the device in the presence of noise, temperature changes or other perturbations. In view of the earlier presented data on the temperature sensitivity of high mobility InSb films, it is desirable to extend the linear range of v_h vs. I by pulse driving the Hall gener-

ator and to improve the thermal dissipation of heat from the film by conduction. It is not considered desirable to dope the film with n-type impurities since this reduces the field sensitivity of the Hall generator.

The construction of a pulsed experimental multiplier is shown in Fig. 7. The magnetomotive force is supplied by the current i_c through the solenoid mounted in the ferrite cup core shown. The latter also contains the Hall plate mounted on its center posts with the leads connected in such a manner as to minimize inductive coupling. Thermal contact is provided between the film and the ferrite by means of silicon grease. Provided that B is below saturation, hysteresis effects are negligible and $R_h(T)$ is considered independent of temperature:

$$v_h = k\, i_c i_h \tag{13}$$

where k is constant and a parameter of the magnetic circuit and the Hall plate. Pulsed operation of such a device requires the synchronization of i_h and i_c and the correction of the pulse unbalance potential in addition to the earlier mentioned reduction of the dc misalignment potential. A suitable circuit for pulse driving the multiplier is shown in Fig. 8 and the results obtained with it are shown in Fig. 9. For either dc or pulsed operation the multiplier output is proportional to the product of i_h and i_c with an accuracy better than ± 1%. Within the range of i_h shown in Fig. 9, thermal drift is negligible although higher values of i_h produce non-systematic fluctuations in the null balance of the circuit and might have to be surpressed by thermal compensation of the circuit.

Other analog computer functions[12] such as division may also be performed by means of Hall generators. Figure 10 shows a circuit using a Hall multiplier whose output v_h is compared against an arbitrary input signal v_i at the input of a differential amplifier. The amplified difference $(v_h - v_i)$ determines the output current and hence the light intensity of the lamp *l*. The latter controls the resistance of the photoresistor r and hence the current i_h of the Hall generator. Thus the electrooptic transducer composed of *l* and r constitute a feedback loop around the differential amplifier and also isolate the input from the output circuit of the Hall plate.

For a finite misalignment potential of the Hall generator, $v_\mu = \beta i_h$ where β = constant, the Hall current with the circuit in the quiescent state is given by:

$$i_h = \frac{v_i/(k\, i_c)}{1 - \beta/(k\, i_c)} \tag{14}$$

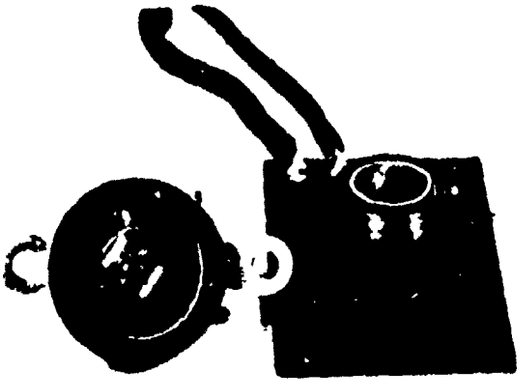


Figure 7. Analog multiplier showing Hall plate mounted on ferrite cup-core, lead arrangement for minimizing inductive pickup, and magnetizing solenoid. Silicon grease is introduced in gap between Hall plate and ferrite center post for decreasing current induced heating of film.

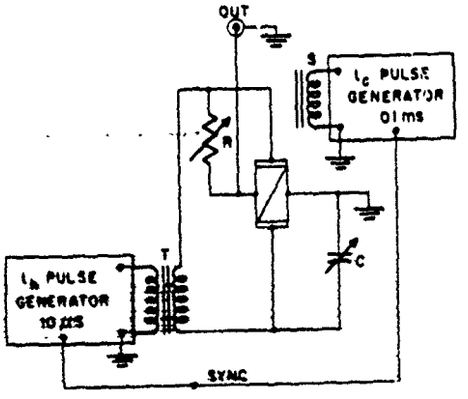


Figure 8. Circuit used for pulsed operation of thin film Hall generators. Pulse transformer T isolates Hall plate from generator supplying pulse current i_h. The mmf supplied by pulsed current i_c through solenoid S is synchronized with i_h. Resistor R and variable capacitor C are used to balance misalignment potential.

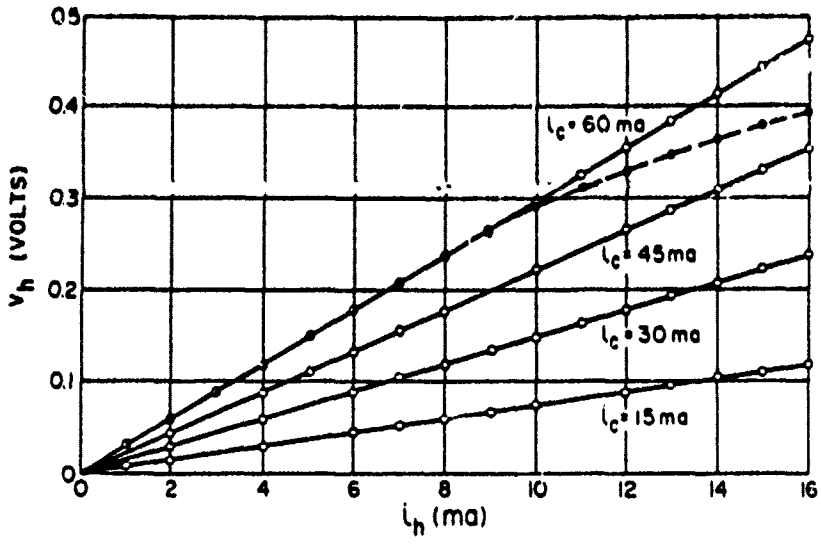


Figure 9. Hall voltage output vs. drive current of pulsed multiplier for fixed values of i_c. Pulse data is shown by open circles. Filled dots show behavior of multiplier for both i_c and i_h driven from dc current sources.

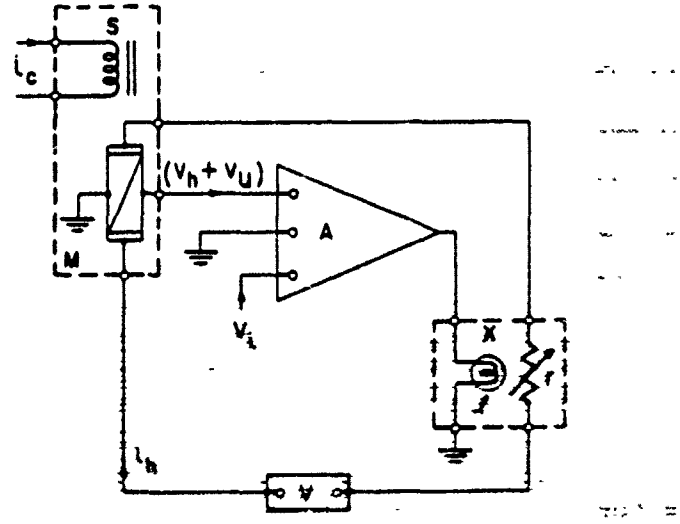


Figure 10. Analog divider based on the Hall effect. The analog multiplier M whose output is amplified by the differential amplifier A has an input current i_h (derived from voltage source V), controlled by electrooptic transducer X in the negative feedback loop of A.

The Hall current will be proportional to the ratio $(v_l/k\ i_c)$ for $(\beta/k\ i_c) \gg 1$ hence the accuracy of division depends upon reducing the misalignment voltage to a minimum and increasing the parameter k of the multiplier (equation 2) to its maximum.

A dc multiplier, similar to that shown in Fig. 7, was incorporated into a ratio computer built in accordance with the diagram of Fig. 10. It was found that over a 4:1 range of i_c, the Hall current i_h is indeed proportional to the ratio $(v_l/k\ i_c)$ to within 2%. This illustrates another potential use of Hall generators, although a disadvantage of this device is its lack of long term stability because of inherent thermal drift and hysteresis connected with this particular feedback transducer.

Hall generator multipliers may also be used for building a simple analog square root computer. The construction and operation of such a device are detailed in Fig. 11. The apparatus enclosed by the dashed square labeled HG, comprises a commercially available Hall generator plate (Model BH-200, Bell Inc. 1356 Norton Ave., Columbus, Ohio) mounted on the center post of a ferrite cup core (Ferroxcube Corp. of America, Type K3-001-01). Within the cup core a magnetic field is generated perpendicular to the plane of the Hall plate by a solenoid also shown in the enclosed rectangle to be in series with the drive-current electrodes of the Hall plate. The Hall voltage output v_h, of the Hall plate is proportional to the product of the magnetic induction B and the drive current i_h or: $v_h = \alpha B i_h$ where α is a material parameter to be considered here as a constant. The magnetic induction is a function of a parameter β, which includes the permeability and geometry of the core and of the magnetizing current which in this case is also the Hall current i.e., $B = \beta i_h$ and therefore:

$$v_h = \alpha\beta i_h^2 \qquad (15)$$

The Hall voltage v_h, proportional to the square of the drive current, is then applied to one of the input ports of the differential amplifier A. The other input port receives the input signal voltage v_o. The difference $(v_o - v_h)$ amplified by the gain of the differential amplifier, drives the grounded emitter power amplifier (using a 2N652 transistor.) The output current of the latter stage drives an incandescent lamp whose light intensity output determines the corresponding resistance of a photoresistor. The purpose of this electro-optic transducer in this circuit is to isolate the Hall plate input from the output and thus prevent severe loading of the amplifier A.

The 2N174 power transistor in a common emitter configuration controls the current i_h obtained from a 10 volt voltage source.

Let $v_o = 0$, the Hall current is then i_h = 1 ma and v_h is substantially zero. Let a potential v_o be applied to the input of amplifier A. The amplified output signal drives the base of the

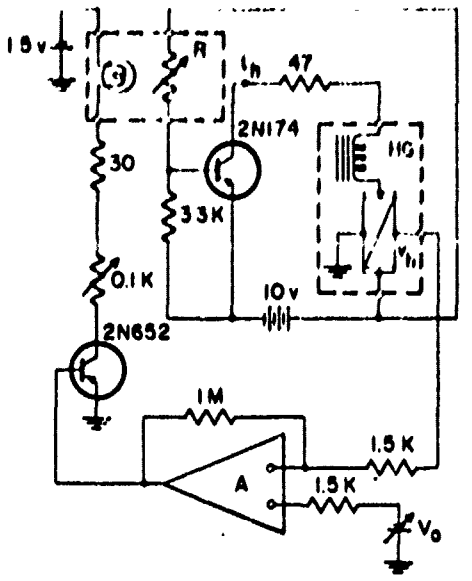


Figure 11. Circuit diagram of square root analog computer. H·G is a squaring multiplier using a bulk InAs Hall plate in series with a solenoid generating a mmf by means of current i_h.

The "Raysistor" R, is an electrooptic transducer completing the feedback loop around differential amplifier A whose input ports are connected to the Hall voltage v_h and to the arbitrary signal v_o.

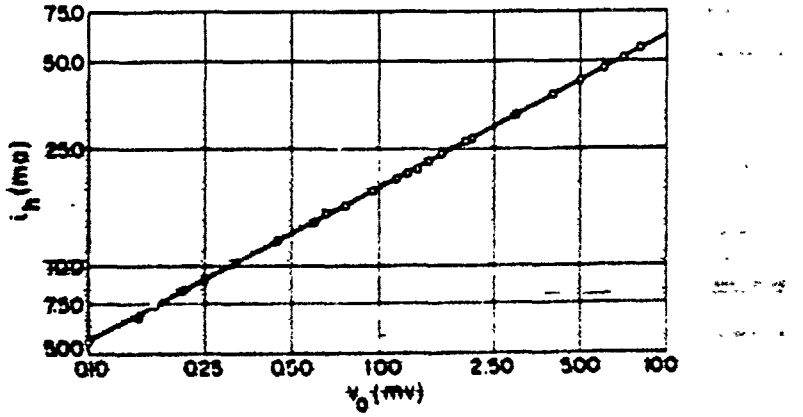


Figure 12. Results obtained with square root circuit of Fig. 11. The Hall current is shown to be proportional to the square root of the input signal v_o over two decades.

2N652 causing a current to flow through the raysistor lamp and sharply decreasing the resistance of the photoresistor. This causes a decrease in the bias applied to the base of the 2N174 power transistor and consequently causes a current i_h to flow through the solenoid winding in series with the Hall plate. Thus a Hall voltage is generated and appears at the other input terminal of amplifier A. The current i_h and hence the Hall voltage v_h, increase up to the time when $v_h = v_o$ when they reach a steady state condition hence:

$$v_o = \alpha\beta i_h^2 \tag{16}$$

Thus the drive current is proportional to the square root of the input potential v_o:

$$i_h = (v_o/\alpha\beta)^{\frac{1}{2}} \tag{17}$$

Figure 12 shows the preliminary results obtained with the circuit of Fig. 11. A log-log plot of i_h vs. v_o is seen to have a slope of $(\frac{1}{2})$ in agreement with equation (5) over two decades of the input potential with an accuracy better than 1.5%. Experiments presently in progress are designed to extend this dynamic range, to improve the accuracy of computation and to extend the operation to alternating frequencies as well as the dc conditions shown here.

The advantage of this method for computing the square root of an electrical input signal rests to a large extent upon the squaring circuit, i.e. the use of the Hall effect in an intermetallic semiconductor such as indium antimonide for obtaining an output which is proportional to the input current squared. The linearity of the Hall effect squaring circuit is particularly suited for obtaining a wide dynamic range and its ruggedness and simplicity and stability is particularly advantageous for portable use. Suppose that the voltage source V = 10 volts in Fig. 12 is a source of a.c. potential. The drive current i_h will also be a.c. Let v_o be an a.c. voltage of the same frequency and in phase with i_h, then A may be an a.c. differential amplifier and the electrooptic transducer may then be replaced by a transformer and still the essential input to output isolation may be achieved. A square root computer may be designed for wideband frequency response provided that due account is taken of synchronization between v_h and v_o.

An ionized gas lamp may be used instead of the filamentary lamp shown in Fig. 2 for the raysistor. Preliminary experiments also show the effectiveness of electroluminescent panel, photoconductor type device is useful instead of the raysistor although it is much less sensitive.

The differential amplifier A may be replaced by a single ended d.c. amplifier provided that the input signal v_o is placed in

series with the output signal v_H.

Thin film Hall generators may, of course, be used in building such an analog square root computer provided that due account is taken of the limited current dissipation of the film and the required mmf for the core. On the other hand the introduction of a suitable current amplifier in the division circuit illustrated in Fig. 12 permits the utilization of standard commercial Hall generators for building an analog ratio computer based upon the Hall effect.

DISCUSSION

Present techniques of evaporation do not, as yet, allow a satisfactory control of the galvanomagnetic properties of film type Hall generators. In the range between 10^{-5} and 3×10^{-4} cm, the virtual electron mobility increases monotonically with thickness by nearly one order of magnitude. Thick films have definitely larger size crystallite grains. The increase in μ is thus consistent with the data of Guenther and Freller[13] who showed that large crystallites cause an increase in μ for either InSb or InAs films. Films thicker than 5×10^{-5} cm also exhibit a large random spread in μ and σ for Hall generators prepared under nominally identical conditions. From measurements made on 20 films, 10^{-4} cm to 2×10^{-4} cm thick in their extrinsic temperature region, a 2:1 spread in μ and σ was found to be due to impurities incorporated during the deposition process. Thin films (less than 10^{-4} cm) show smaller variations (of the order of 50%) in these parameters, but also consistently lower mobilities. Thus both the degree of crystallinity and the impurity content needs to be considered in improving the sensitivity of film Hall generators.

While the experimental data presented here is concerned with InSb films, this compound having been chosen because of its high electron mobility, other intermetallic binary semiconductors such as InAs or ternary semiconductors using In, P and either Sb or As might be more suitable for application requiring a smaller dependence upon temperature at the expense of a lower μ. Other III-V semiconductors with a wider band gap than InSb also have a higher thermal conductivity and a smaller coefficient of thermal expansion consequently they might prove to be more suitable for Hall effect applications[13] where maximum field sensitivity is not the primary consideration.

The substrates used for these films are relatively thick. Glass has a disadvantage in being a poor thermal conductor, it is, however, a stable and chemically inert material at the required substrate temperatures of the order of 300°C. Conceivably, other materials having in addition a thermal expansion coefficient closer to that of the semiconductor films, might be superior to glass. Preferably the InSb film should be deposited directly upon a ferrite base or upon an inert insulating film first deposited upon the ferrite substrate so as to increase the effective permeability by decreasing the gap spacing of the magnetic circuit.

ACKNOWLEDGEMENT

This work is supported by the Bureau of Naval Weapons under a continuing program of research on the properties and applications of thin film devices based upon the Hall effect. The writer is indebted to R. Potter, Infrared Division, U.S. Naval Ordnance Laboratory, Corona, for evaporating and processing the InSb films, and to his associates A. Clawson and D. Collins for considerable aid in the experimental phase of their investigations.

BIBLIOGRAPHY

1. K. G. Guenther, "Evaporated Films of III-V Semiconducting Compounds", Z. Naturf., Vol. 13a, pp 1081-1089 (1958).

2. W. Haenlein and K. G. Guenther, "Preparation and Properties of Multicomponent Films by Means of Evaporation in Vacuum", Adv. in Vacuum Science and Tech., Vol. II, pp 727-733 (1960).

3. R. F. Potter and G. G. Kretschmar, "Optical Properties of Evaporated InSb Films", J. Opt. Soc. Am., Vol. 51, pp 693-696 (1961).

4. S. Tolansky,"Mutiple-Beam Interferometry of Surfaces and Films", Oxford: Clarendon Press (1948).

5. I. Isenberg, B. R. Russell and R. F. Greene, "Improved Methods of Measuring Hall Coefficients", Rev. Sci. Instr., Vol. 19, pp 685-692 (1948).

6. E. H. Putley, "The Hall Effect and Related Phenomena", Butterworth Scientific Publications Ltd., London (1960).

7. V. V. Galavanov, "On the Temperature Dependence of Hall Generators", Soviet Phys. - Solid State 2, 55 (1960); Fizika Tverdogo Tela 2, 62 (1960).

8. H. Welker and H. Weiss, "Group III-V Compounds", Solid State Physics Advances in Research and Applications, Vol. 3, Academic Press, Inc., New York (1957).

9. H. H. Wieder, "Hall Effect Stroboscope and Noise Discriminator", Rev. Sci. Instr., Vol. 33, pp 64-70 (1962).

10. H. H. Wieder, "Sampling Magnetometer Based on the Hall Effect", J. Appl. Phys. (Suppl.), Vol. 33, pp 1278-1279 (1962).

11. A. R. Clawson, Betty Knorr and H. H. Wieder, "Bibliography of the Hall Effect, Theory, Design and Applications", Naval Ordnance Laboratory, Corona, Report 7233, September (1962).

12. H. H. Wieder, "Analog Division Based on the Hall Effect", Rev. Sci. Instr. 34, April 1963 (in press).

13. K. G. Guenther and A. Fuller, "Properties of Evaporated InSb and InAs Films", Z. Naturf., Vol. 16a, pp 279-283 (1961).

EPITAXY OF PbS, PbTe, AND SnTe

Jay N. Zemel
U. S. Naval Ordnance Laboratory
White Oak, Silver Spring, Maryland

In the language of present day technology, epitaxy refers to the oriented overgrowth of a particular substance such as germanium or silicon on a parent crystal. One particularly well-known example is the epitaxial transistor. Oriented overgrowths of many materials can be formed on a variety of crystalline substrates and for many years this was a well-known scientific curiosity.[1,2] But, as so often happens, yesterday's curiosity is tomorrow's technology, especially in solid state phenomena. To make this transition requires a good deal of understanding of fundamental processes and today we are on the road to this understanding.

Our particular interest in the materials PbS, PbTe and SnTe dates back to the days when the lead salts were practically the only photoconductive infrared detectors. While the photoconductive properties of these materials are still poorly understood, our main concern at the Naval Ordnance Laboratory has been more with the band structure and fundamental processes than with direct applications of these materials. Particular stress placed on producing high quality bulk single crystals[3] had been reasonably successful but work still continued on photoconductive films of polycrystalline PbS. One particular phase of this work dealt with the mobility of holes in these p-type films.[4] It was noticed that the mobility was not dependent on the carrier concentration for a set of films having the same thickness. It was

decided to investigate this question at greater length using thin single crystals. Since it is completely impractical to remove material from a thick crystal of these soft and friable materials to form one with a thickness in the range 1000 A to 50,000 A, our only choice was to seek a way of growing these films. It was well-known that PbS would grow on NaCl but what we did not know was whether the resulting films would resemble the bulk material or the usual evaporated films.[1,2] The results of our investigations established quite dramatically that the film properties are much closer to the bulk than they are to those of polycrystalline evaporated films.[5,6] It is this characteristic which raises some interesting, indeed exciting, prospects for the study of compound semiconductor materials.

Let us first establish our case with some direct experimental evidence. To begin with, we produce the films by directly evaporating from a quartz oven onto a heated alkali halide substrate. The vacuum is not extraordinarily good, being about 5×10^{-6} to 10^{-5} mm of Hg during the evaporation. This does not seem to have too deleterious an effect on the properties of the films. The substrate is maintained at a uniform temperature during the deposition ranging from 250°C - 325°C. In figure 1 is a back reflection Laue pattern for the first of the materials studied, PbS. As can be seen, this pattern is clean and clearly indicative of the single crystal character of the film. Additional x-ray evidence of this is obtained from Bragg diffractometer recordings shown in figure 2. An iron target was used and the Kα iron doublet lines are clearly resolved by the PbS. While this information is important and significant, the electrical properties are generally a far more sensitive test for the presence of imperfections and dislocations in a material than either optical or x-ray measurements.

In figure 3 we show the resistivity of several films as functions of reciprocal temperature. The temperature range encompassed by these measurements is from 300°K to 77°K; that is from room temperature down to liquid nitrogen temperatures. The resistivity is characteristic of bulk PbS in that it decreases rapidly with decreasing temperature. This is in sharp contrast to the behavior of photoconductive films of PbS prepared

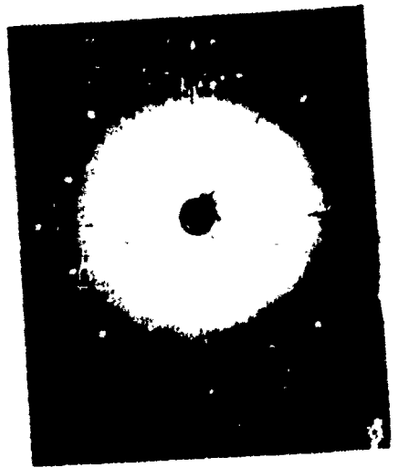


Fig. 1 Back reflection Laue pattern of PbS epitaxial film

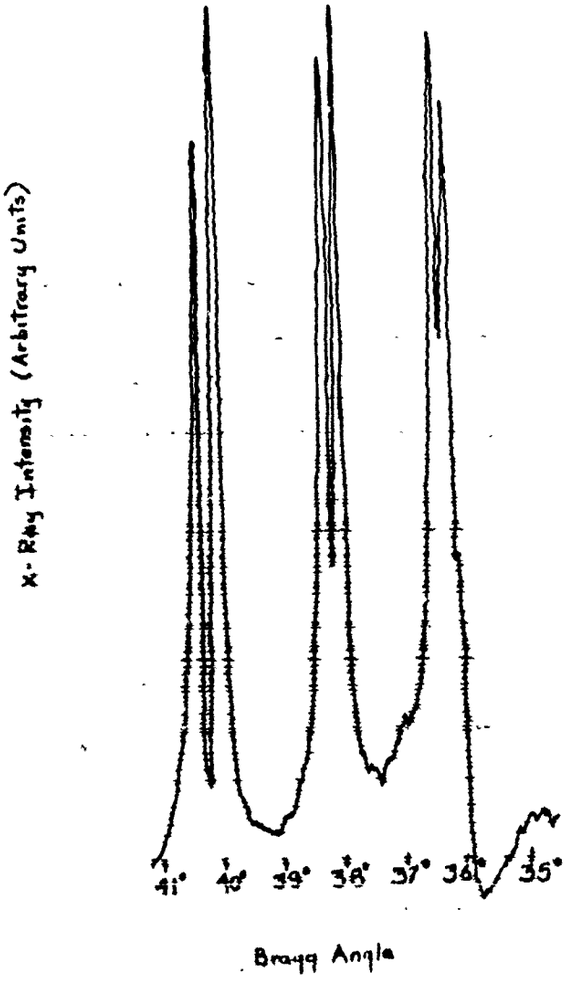


Fig. 2 Tracing of a Bragg diffractometer recording of a PbS epitaxial film. The iron $K\alpha$ doublet is clearly resolved indicating good equipment alignment and good crystal perfection.

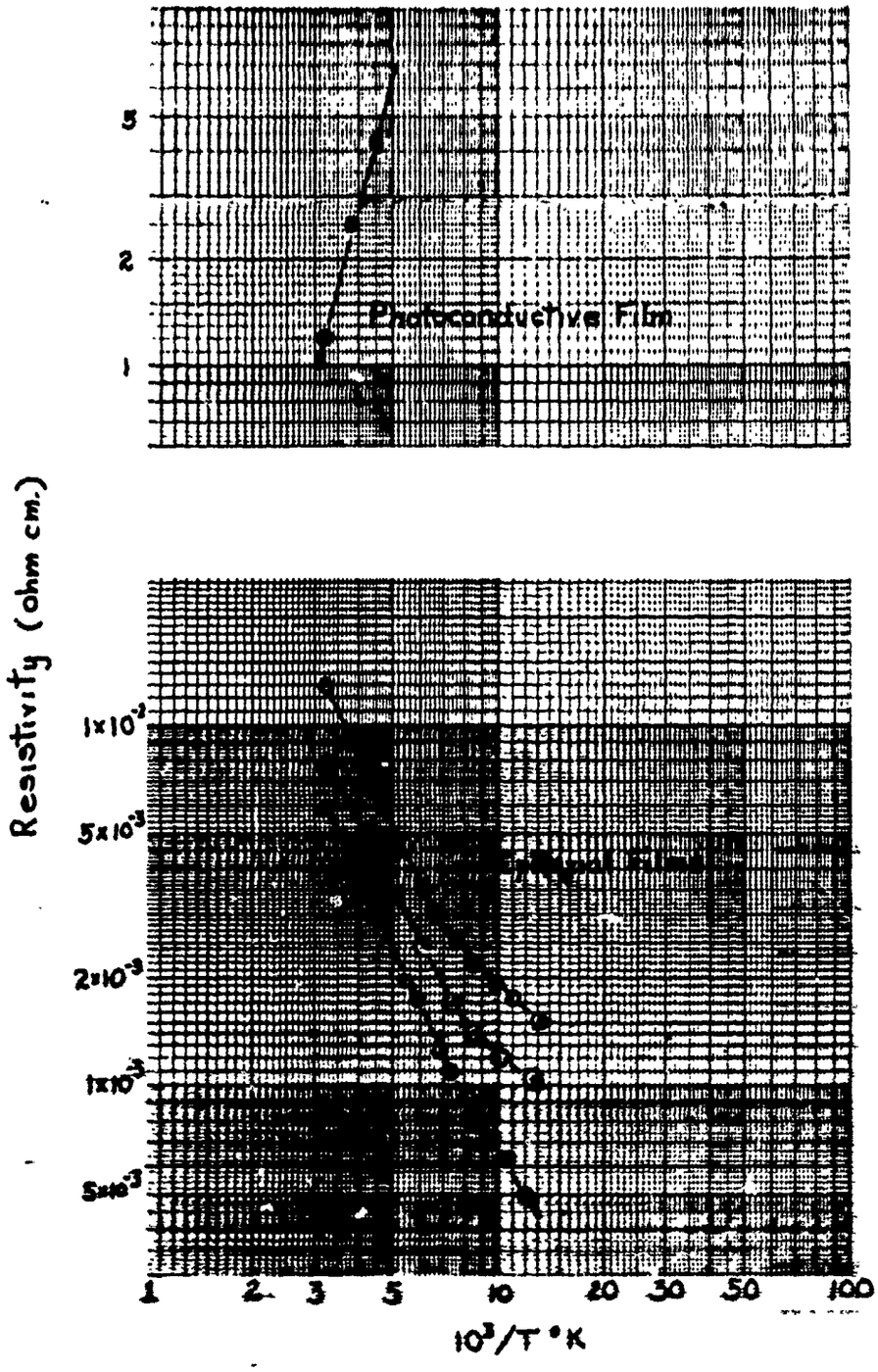


Fig. 3 Resistivity of PbS versus reciprocal absolute temperature from room temperature to liquid nitrogen temperature

by chemical deposition. In order to see clearly the differences between conventionally produced thin films and epitaxial films, we will describe the PbS photodetector's properties and contrast them with the epitaxial films. In the chemically deposited films, the resistivity rises very sharply. This increase is due to a decrease in the carrier density as the temperature decreases. The origin of this effect is associated with the structure of the film and the surface states present in the film. The carrier density in the epitaxial films on the other hand, is quite characteristic of n-type bulk as shown by the temperature dependence of the Hall coefficient shown in figure 4. It indicates very little change in carrier concentration with temperature. The Hall mobility is also markedly different in the two materials, that is chemically deposited and epitaxial PbS. This is shown in figure 5. The thickness of the films are essentially the same in the two curves. It is clear that the small crystallites composing the chemically deposited films have radically different properties than the thin single crystal films and these properties we feel depend strongly on surface rather than bulk phenomena.

This difference serves as a point of departure for our research program on the mechanism of photoconductivity in the lead salts, PbS, PbSe and PbTe. The U. S. Navy, and the Defense Department in general, has a substantial interest in the PbS-type detector. Not only does it cover an important part of the infrared spectrum, it is also extremely convenient for application to defense problems. The difficulties with its manufacture are intimately tied up with the mechanism of photosensitivity and, just as the sources of difficulty are rarely understood, so too are the reasons why they disappear. With the epitaxial film we start afresh with the relatively simple, well-defined properties of a single crystal. Eventually, it may be possible to follow in detail the onset of photoconductivity in these films as they are treated with various bulk dopants and surface reactants.

Before that, we must understand the properties of bulk PbS and epitaxial PbS quantitatively. So far we have been primarily concerned with the electrical transport properties of PbS. Let us consider the optical properties of these films. This is particularly simple

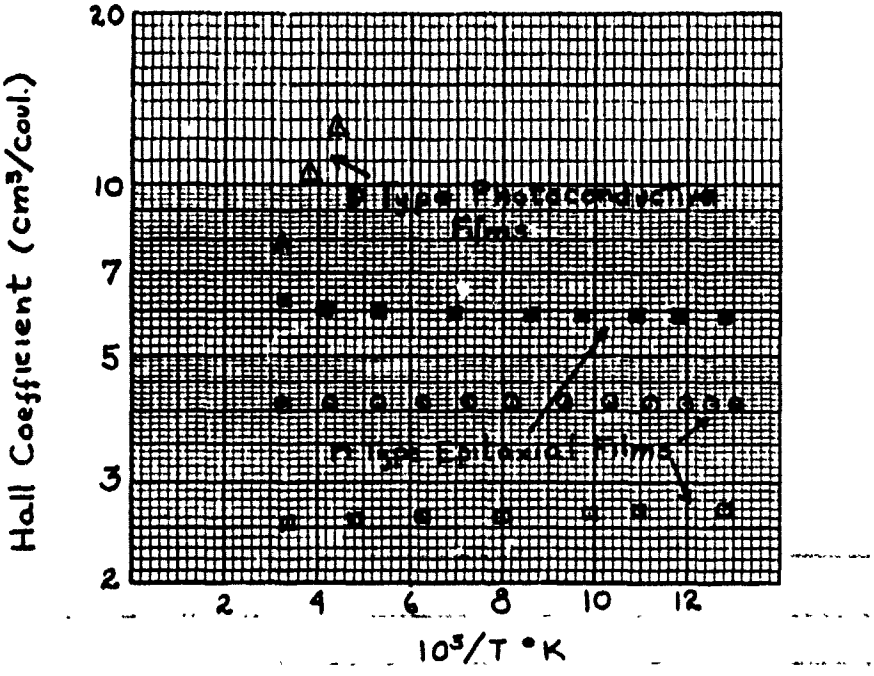


Fig. 4 Hall coefficient of PbS versus reciprocal absolute temperature from room temperature to liquid nitrogen temperature

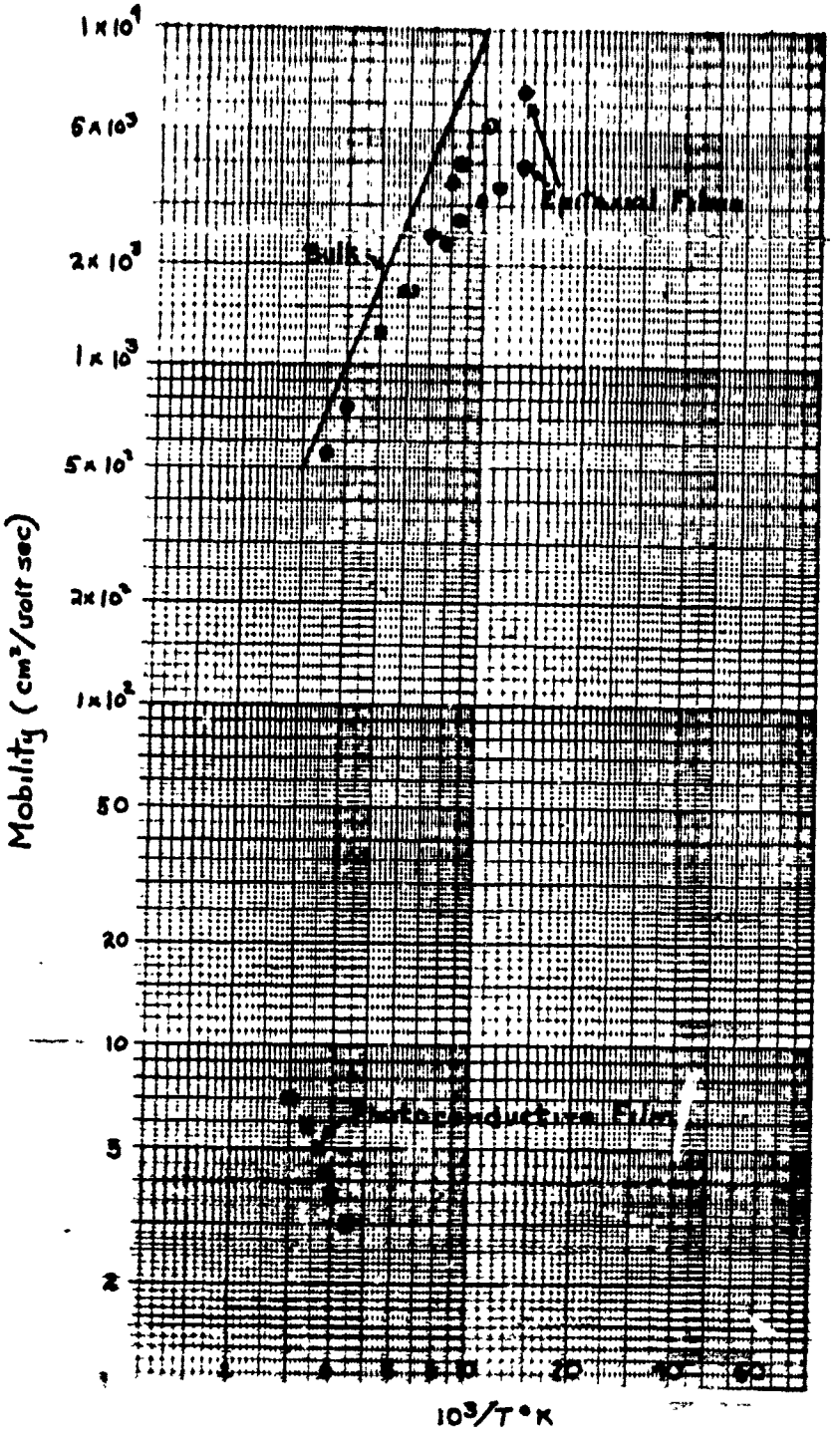


Fig. 5 Mobility of PbS versus reciprocal absolute temperature from room temperature to liquid nitrogen temperature

and convenient to do since the films tend to be extraordinarily uniform in thickness, variations of which rarely exceed a few percent over distances of 5 mm. Thus the interference fringes are sharp and well defined. It is possible to identify the first order interference fringe. This is shown clearly in figure 6. The films are weighed after they are removed from the substrate by the simple device of dissolving the alkali halide away in water. Using the x-ray value for the density, the thickness is readily obtained from the mass of the film. Having the thickness, the order, and the wavelengths of the maxima and minima of the interference patterns, one can readily determine the index of refraction from the relationship

$$m\lambda = 2\mu d$$

where m is the order number, λ the wavelength, μ the index of refraction and d the film thickness. The index has been measured by Schoolar and Zemel[6] by this method from 25 microns through the absorption edge (3 microns) to 1.5 microns. The results are shown in figure 7. What is particularly interesting is that the dispersion peak in the vicinity of the band edge is in excellent agreement with the theoretical predictions of Stern.[7,8a] It is now fairly apparent that the films are well suited for optical investigations and their initial impact will be in this important scientific area.

The next question is whether other materials can be grown in the same fashion as PbS. Keep in mind that our vacuum conditions are far from the best and while PbS did not seem to be especially sensitive to pressures in the 10^{-6} mm of Hg range, there was no reason to believe that we would be as fortunate with other materials. We next tried two materials of interest to our research effort at the Naval Ordnance Laboratory, PbTe and SnTe. The results are best shown in figures 8, 9, and 10.

In figure 8 we have the Laue pattern for PbTe. As was the case with the PbS, the spots are sharp and well defined with no apparent "ringing" due to misorientation. The results of Bragg diffractometer studies are equally encouraging. Let us consider the electrical

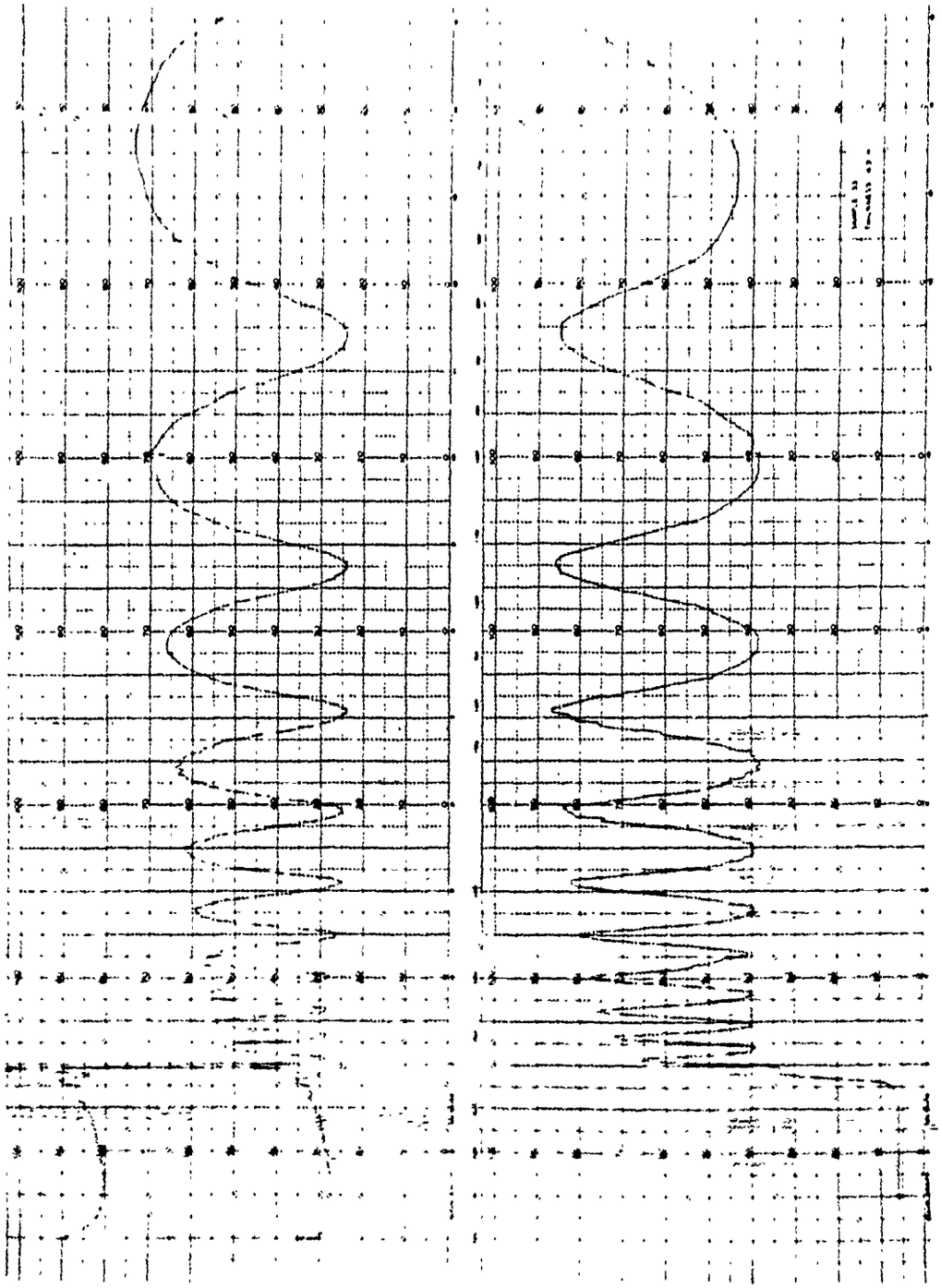


Fig. 6 Reflectivity and transmissivity of a 4.3 micron thick PbS epitaxial film on rock salt

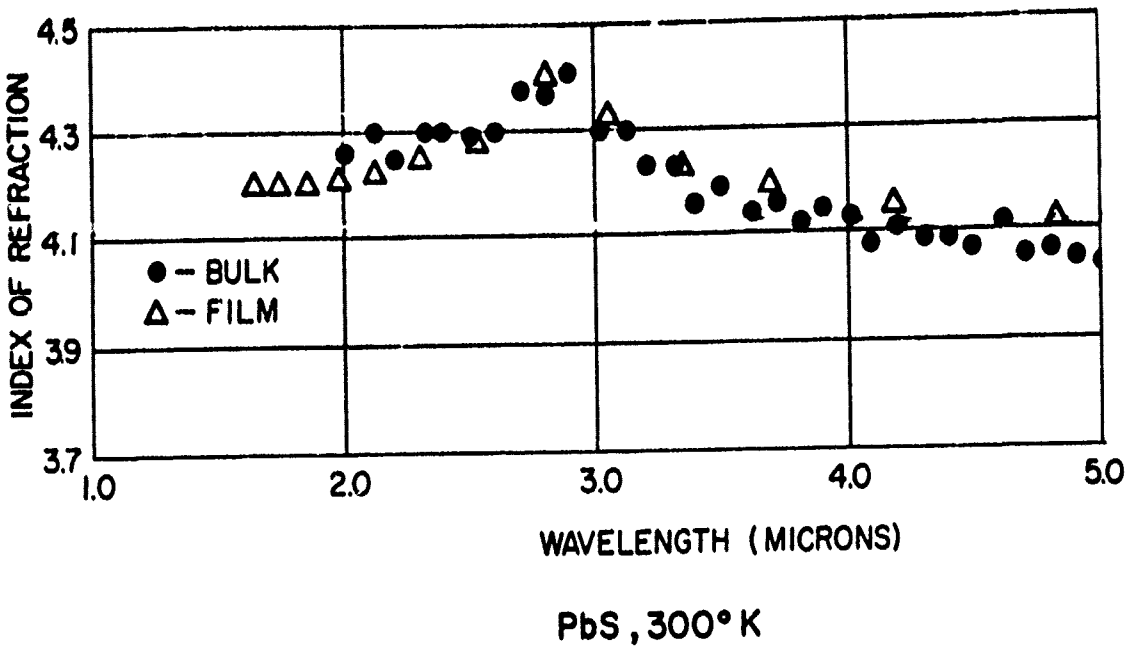


Fig. 7 Index of refraction of PbS as a function of wavelength

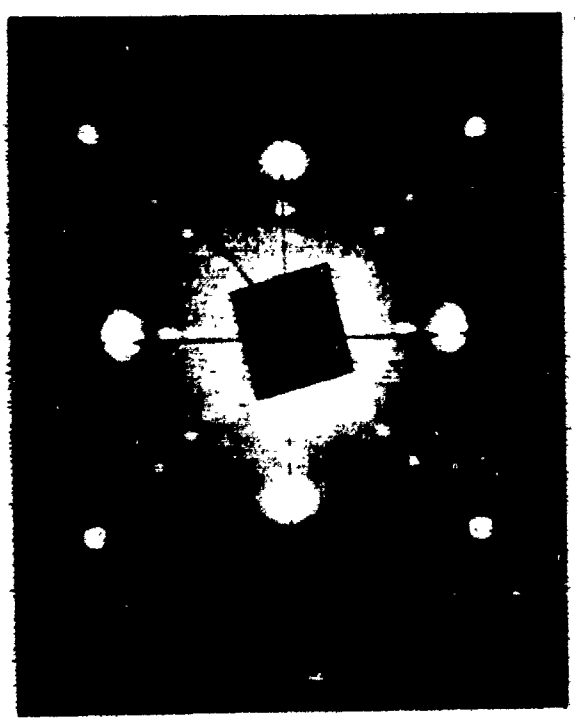


Fig. 8 Back reflection Laue pattern of PbTe epitaxial film

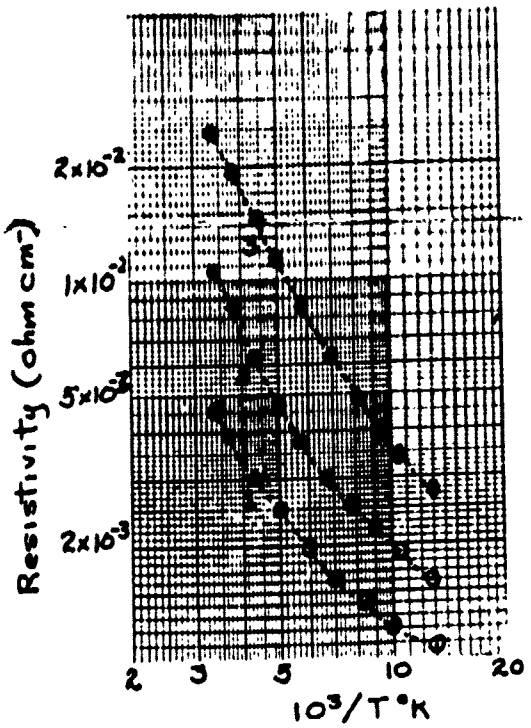


Fig. 9a

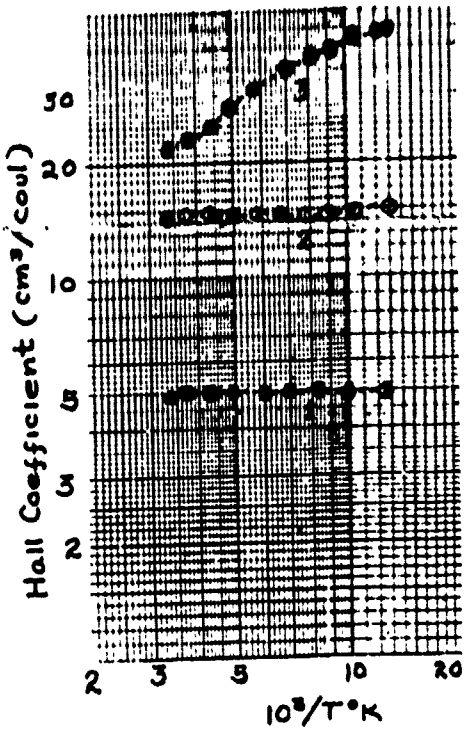


Fig. 9b

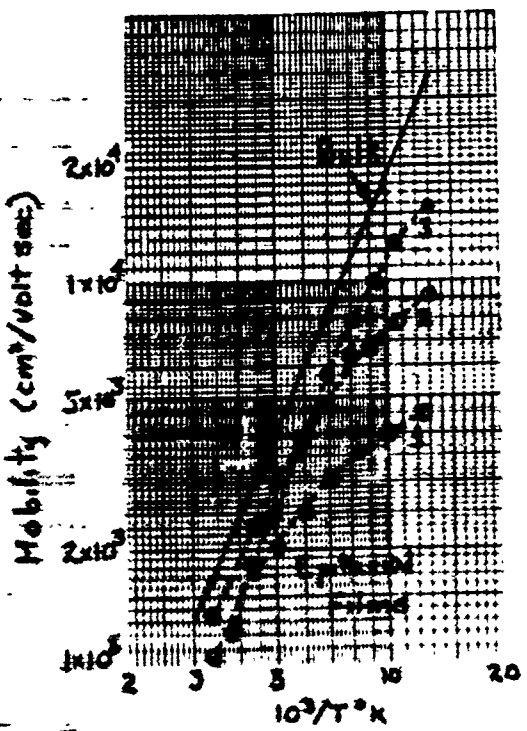


Fig. 9c

Fig. 9 Electrical properties of epitaxial films of PbTe as functions of ~~reciprocal~~ absolute temperature from room temperature to 77°K (a) Resistivity, (b) Hall coefficient, ~~(c) Hall mobility~~

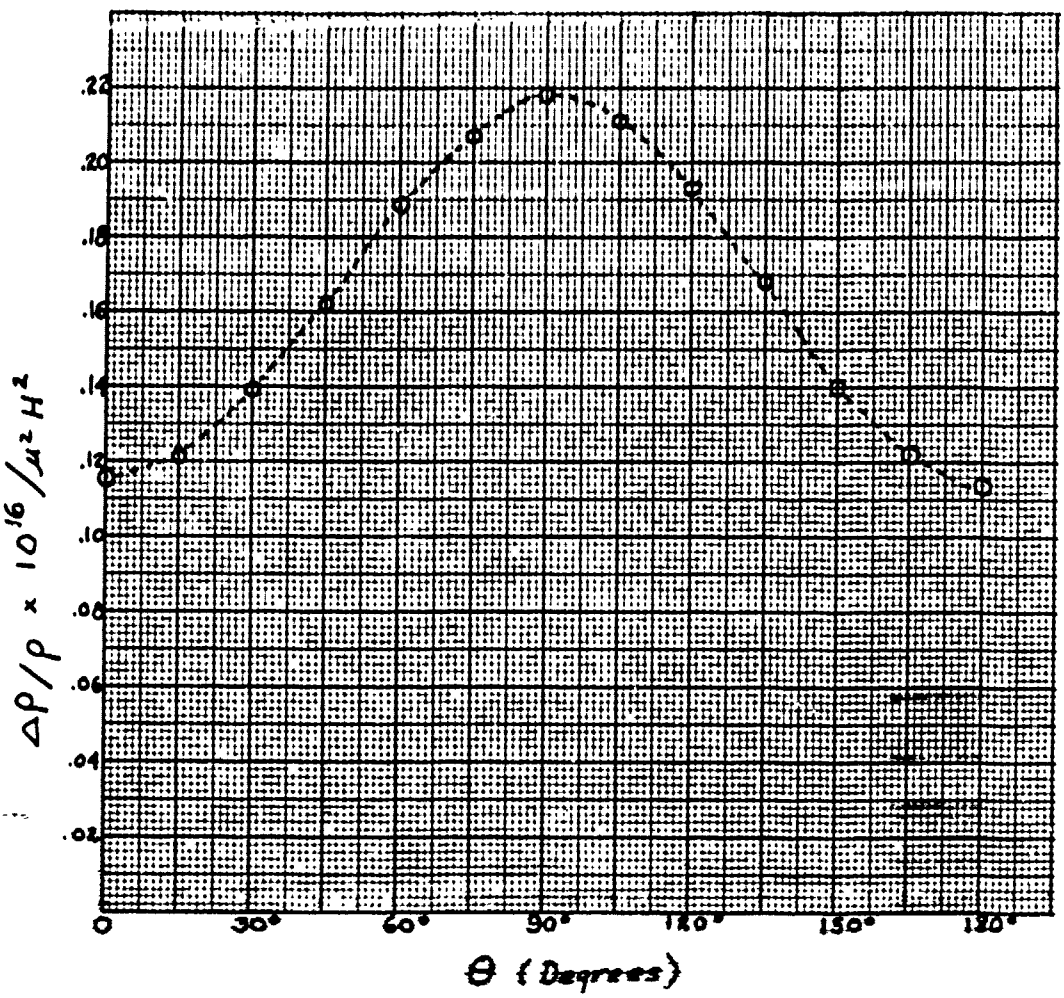


Fig. 10 Angular dependence of the magnetoresistence of PbTe. θ is the angle between the current and the field. μH = 6.18 x 10^7 at 77°K.

properties of PbTe.[8] In figure 9a we show the resistivity for a number of films ranging from 300°K to 77°K. The comparison with the bulk shown here is extremely favorable. Now it is interesting that we have managed to grow both p- and n-type films of PbTe, while we could only grow n-type films of PbS. The behavior of the Hall constant is shown in figure 9b. As can be seen, the Hall constants are fairly large, indicating that some of the films had carrier concentrations as low as $2x10^{+17}$ electrons/cc. Generally, it is uncommon to have PbTe in this carrier concentration range and we are not sure of the interpretation of the temperature dependence of the Hall constant. It could be due to inhomogeneities in the film or a second low mobility conduction band. In either case, we simply are not sure of the explanation. The Hall mobility in figure 9c is excellent evidence of the high quality of the films. As can be seen, the mobility in the best film rises to almost 16,000 cm^2/v sec at 77°K. This high mobility strongly suggests that the density of imperfections is fairly low. One of the more sensitive measures of homogeneity is the magnetoresistance. In figure 10 we show a typical plot of the magnetoresistance at 77°K as a function of the angle between current and the magnetic field in a plane normal to the surface and parallel to the current. This exhibits the expected sinusoidal behavior with the longitudinal magnetoresistance being larger than the transverse. This is in keeping with the observed behavior of bulk n-type PbTe. With the high mobilities found in these films, we would expect that the magnetoresistance would start to show departures from linearity with H^2 when $\mu H>1$. In figure 11 we can clearly see this break at $\mu H = .4$, again in agreement with the bulk. Thus, the electrical properties provide a very solid basis for further study of the PbTe as a thin film. The optical properties are also of interest because of the information to be obtained on the band structure. Since interference patterns are esthetically pleasing, consider figure 12. Here we see almost classically perfect interference fringe. From the maxima and minima of both the transmission and reflection, we can obtain the index of refraction as mentioned earlier. The index is shown in figure 13. Of particular interest is the two peaks in the index. The first peak may be associated with the direct valence-conduction band edge. The second peak is associated

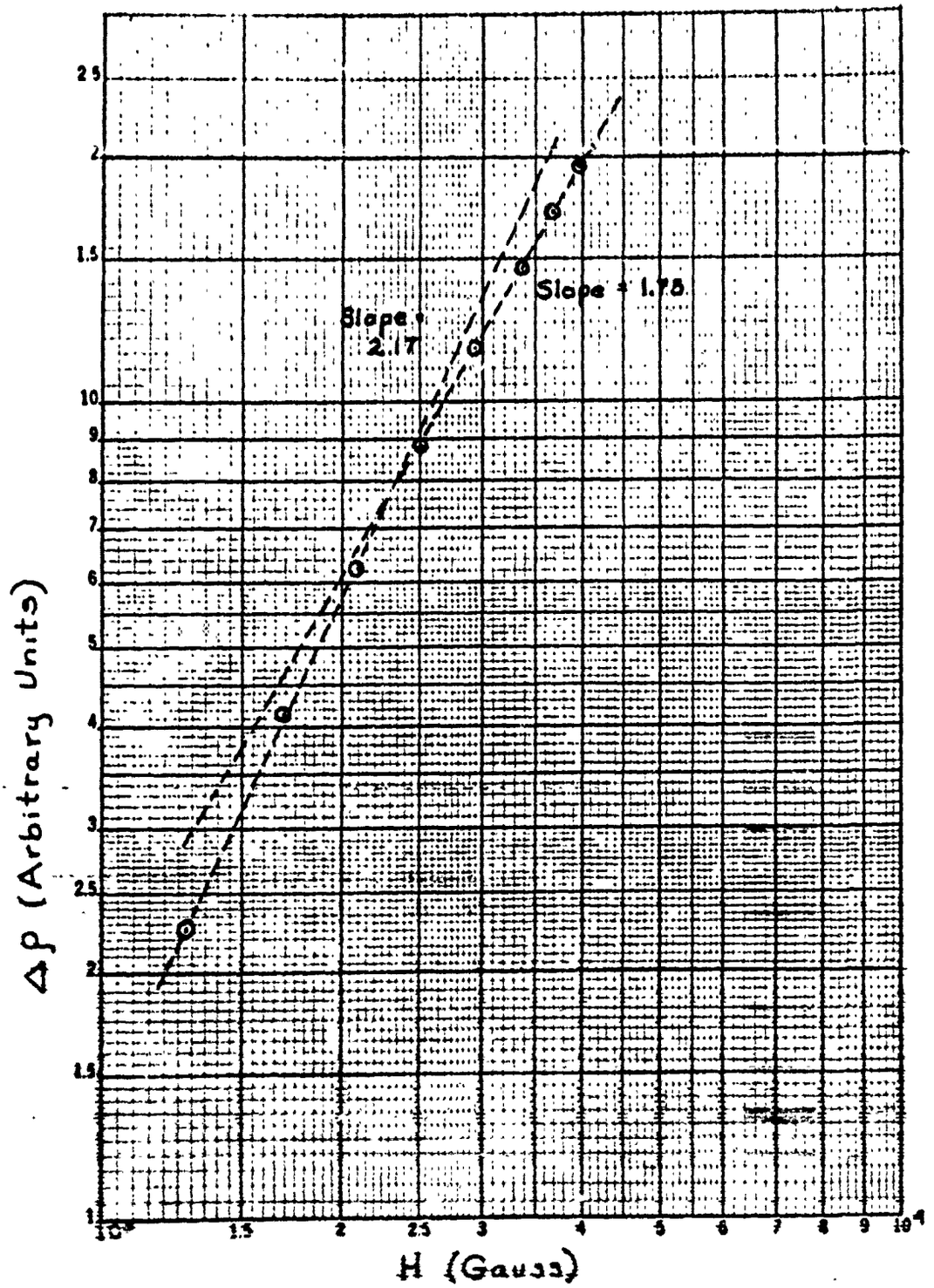


Fig. 11 Magnetic field dependence of the magnetoresistence of PbTe. μ = 16,000 at 77°K.

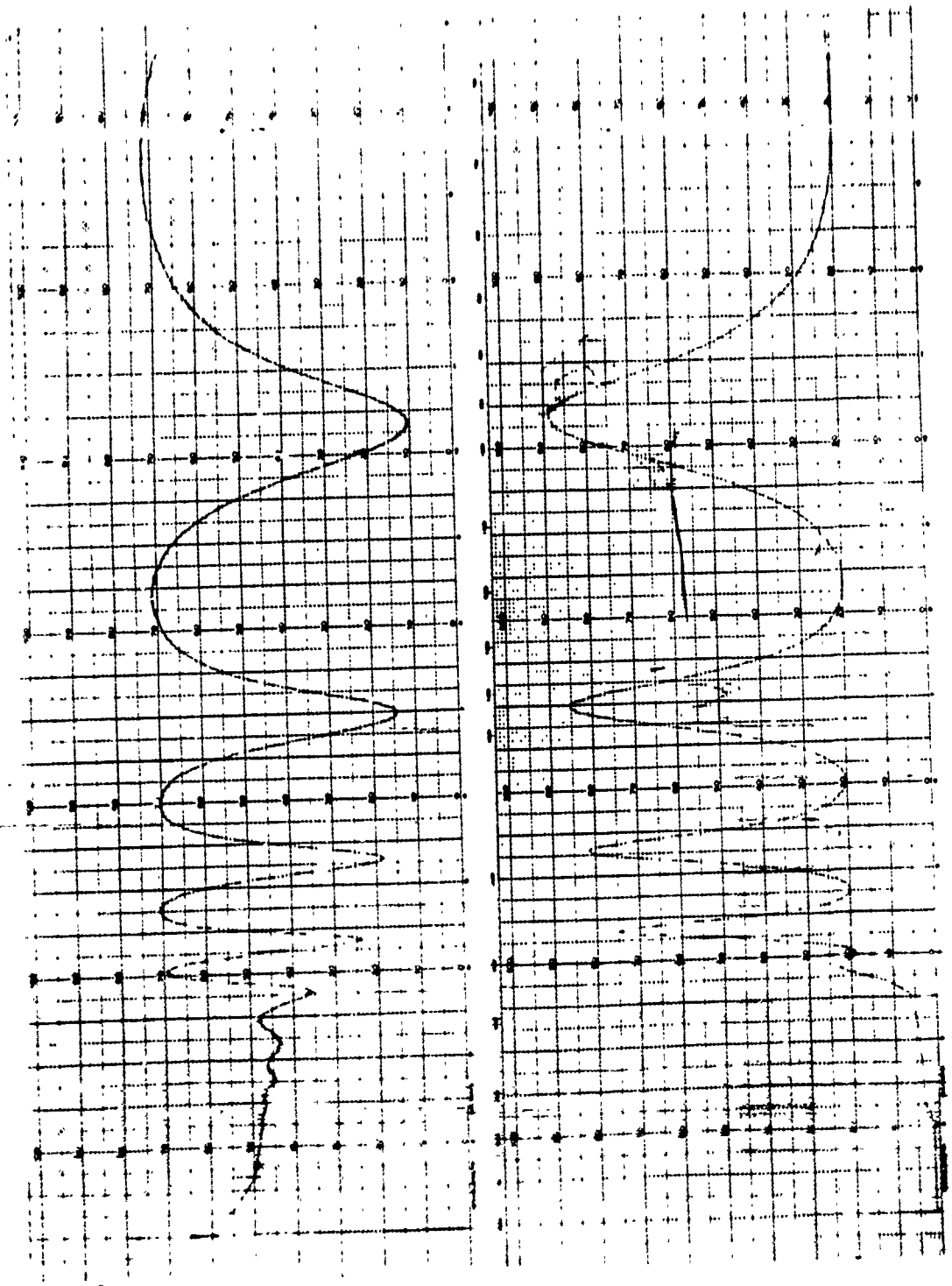


Fig. 12 Reflectivity and transmissivity of a 1.8 micron thick PbTe epitaxial film at room temperature

with a second band and is believed to be a spin-orbit split-off band. Since we are dealing with single crystal films, we do not have the aging problems other investigations have had using polycrystalline films. We have found, experimentally, that the index of a given film does not change with time.

Now SnTe has also been successfully grown as a thin film and the agreement with the bulk is equally impressive. The x-ray, electrical and optical data all agree very well.(8) The big difference between SnTe and the Pb salts is that SnTe has an enormous density of Sn vacancies and the films have carrier concentrations in the $10^{20}/cm^3$ range. This high a carrier concentration will, of course, lead to an enormous free carrier absorption even at very short wavelengths. There is a "window" between 2 and 4 microns as shown in figure 14. Through this window one can see interference fringes but the "visible" region of the "window" is much too narrow for a useful measurement of the index by interference methods. The index must await the development of wedges so that the index can be measured in a conventional fashion.

As can be seen, the data all points in the same direction. This class of materials is extraordinarily well suited for being grown as single crystal films. The unfortunate part of this is that we do not have a good model which accounts for the growth of these films. We can of course, speculate privately but until we have more evidence it would be premature to do so publically.

Let us speculate a bit on the implications and possibilities of this work. In the scientific realm, it points the way of preparing interesting and important high carrier concentration materials in a convenient form. Cyclotron resonance and interband magneto-optic phenomena all become feasible with these films in many materials. Indeed, Mitchell et al. have been examining the interband magneto-optics of PbS and PbTe with these films.(9) These measurements are expected to yield useful information on the band structure of these materials.

But in addition to this, there is a question of materials preparation that this opens up. First of

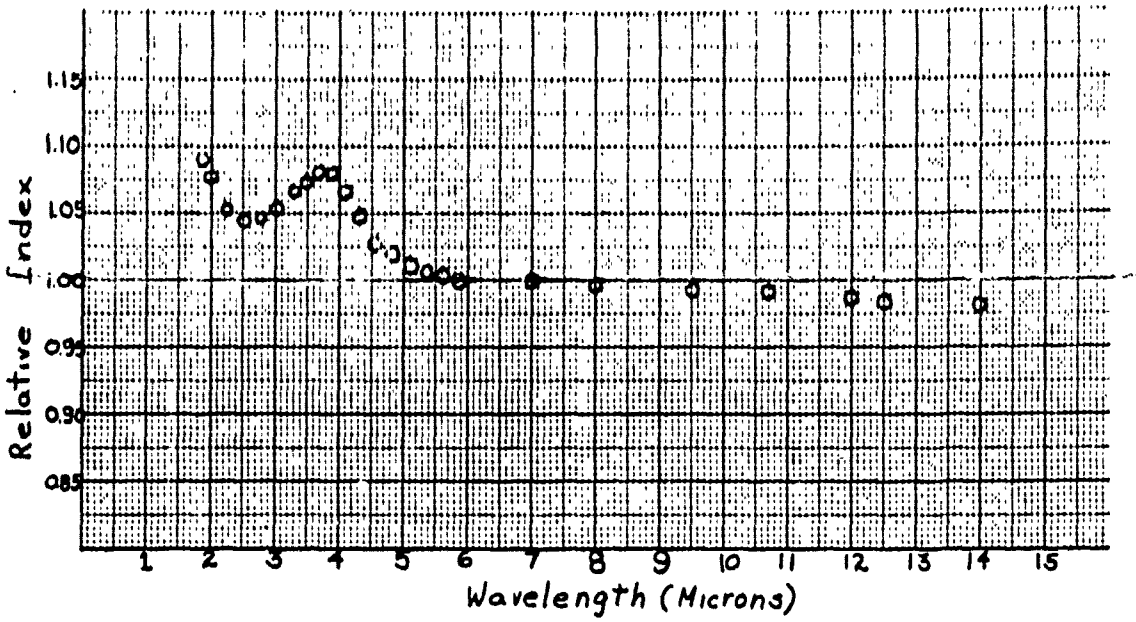


Fig. 13 Index of refraction of PbTe at room temperature as a function of wavelength

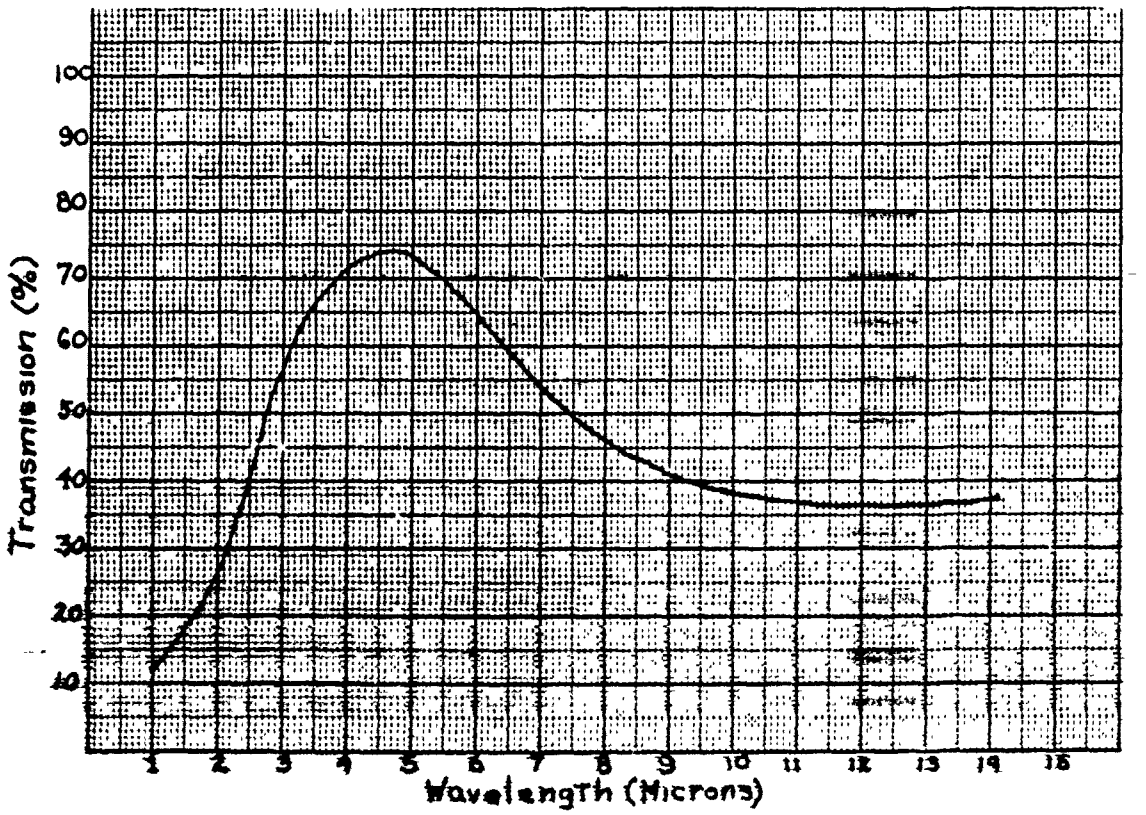


Fig. 14 Transmissivity of a thin film of SnTe of unknown thickness as a function of wavelength

all, thin films are becoming increasingly attractive technologically because of reduction in handling that arises in their preparation. Single crystal films combine, in principle, the best aspects of modern solid state technology. In our procedures we start with crushed polycrystalline material and carry out a fractional distillation, keeping in essence the middle fraction. The material deposited on the substrate forms a single crystal of good quality and in a range of thickness where diffusion will rapidly iron out any inhomogeneities in depth. While we have only touched fleetingly in our own work on the diffusion of dopants into the films, it is enough so that we feel that we can readily study this type of phenomena. Furthermore, it should be equally possible to prepare alloys by simultaneous evaporations. Perhaps the most interesting possibility in this respect is to prepare films where the alloy's composition varies linearly from one end of the film to the other. For survey work, such an arrangement could be invaluable.

What can we conclude from this work? First of all, it appears that good, thin, single crystals of the lead salts can be prepared. Second, many fundamental processes can be advantageously studied using these films. Third, because of the relative simplicity in their preparation, these materials should be investigated more carefully for device application. Finally, consider the original aims and the interim results of this work. We started with our sights leveled on photoconductivity and we find ourselves heavily engaged in band structure, diffusion processes and microelectronics.

Since this paper is a review of the Naval Ordnance Laboratory film program, below are the names of those people who have contributed so much to its current progress; James O. Varela, surface studies on chemically deposited PbS photocells; Richard B. Schoolar, preparations and optical properties; James D. Jensen, electrical measurements. We would also like to acknowledge the assistance of the other members of the Solid State Division who have contributed by their discussions to our work; Jack R. Dixon, Robert A. Allgaier, Richard F. Greene; Bland B. Houston, Jr., and H. Ray Riedl.

REFERENCES

1. D. W. Pashley, Advances in Physics 5, 173 (1956).

2. Structure and Properties of Thin Films, edited by Haugebauer, Newkirk and Vermilyea, published by John Wiley and Sons, New York, 1959.

3. B. B. Houston, Jr, private communication.

4. J. Zemel and J. O. Varela, J. Phys. Chem. Solids 14, 149 (1960).

5. J. D. Jensen and J. Zemel, Bull. Am. Phys. Soc. 6, 437 (1961); R. B. Schoolar, J. D. Jensen and J. N. Zemel, Bull. Am. Phys. Soc. 8, 63 (1963).

6. R. B. Schoolar and J. Zemel, to be published.

7. F. Stern, private communication.

8a. R. Riedl and R. B. Schoolar, submitted to Phys. Rev. Letters.

8. J. D. Jensen and R. B. Schoolar, Bull. Am. Phys. Soc. 8, 198 (1963).

9. D. L. Mitchell, E. D. Palik, J. D. Jensen, R. B. Schoolar and J. N. Zemel, Phys. Letters, to be published.

IN-PILE HALL ELECTRON TRANSPORT PROPERTY MEASUREMENTS ON ZONE-REFINED, P-TYPE SILICON

G. C. Bailey and C. M. Williams
U. S. Naval Research Laboratory
Washington, D. C.

I. INTRODUCTION

Experimental analysis of radiation damage in semiconductors is usually accomplished by observing the post-irradiation temperature dependence of the Hall coefficient, the electrical resistivity, and optical absorption(1-10). Post-irradiation electron spin resonance measurements are also used to study the defect centers produced by particle(11) or gamma ray bombardment. If transient effects exist, they would, of course, be absent by the time the experiments are performed. The present series of experiments represents the first measurements of the Hall coefficient of silicon to be made in-pile, i.e., during neutron irradiation. Also, in contrast to most of the previous investigations which were carried out on pulled silicon having about 10^{18} oxygen atoms/cm^3, the present experiments were made on zone-refined silicon containing less than 10^{16} oxygen atoms/cm^3.(12)

II. EXPERIMENTAL

Samples with dimensions of 1 cm x 0.2 cm x 0.2 cm were cut from zone-refined single crystals of p-type, boron-doped silicon having initial resistivities of 1, 8 and 100 ohm-cm. An electromagnet was designed which was small enough to be used in the NRL reactor but which provided sufficient magnetic field for Hall measurements. The sample and magnet were then lowered into a long tube placed next to the core of the NRL swimming pool reactor, and Hall and conductivity measurements were made remotely by potentiometric methods during irradiation.

Samples B-1, D-8 and G-100* were all irradiated at 100°C with the following exceptions: the temperature of D-8 varied as the ambient reactor temperature and was 79°C initially, 93°C at an integrated fast neutron flux of $90 \times 10^{12} n/cm^2$, and 100°C after $195 \times 10^{12} n/cm^2$. Samples H-100 and J-100 were irradiated at 72°C in the presence of a dry helium atmosphere. Other details of the irradiation procedure are described in an earlier paper(13).

The flux for Samples B-1 and D-8, irradiated at the center-line of the core, was $3.3 \times 10^{10} n/cm^2 sec$ and was measured using activity determinations of monitors of sulfur (2.9 mev threshhold) pellets irradiated with the samples. Samples G-, H-, and J-100 were placed 1 foot, 3.4 feet, and 4 feet up from the core center-line. For these samples, the flux was determined by a separate experiment, taking the center-line core flux to be that as found for Samples B-1 and D-8. Several sulfur pellets were placed at intervals up from the core center-line and the resultant log activity-distance plot gave fluxes for G-, H-, and J-100 of 1.1×10^{10}, 1.1×10^{9}, and 3.6×10^{8} $n/cm^2 sec$, respectively.† The reactor power level for Samples B-1, D-8, and G-100 was 100 kw whereas for H- and J-100, the power was 1 Mw. This way of determining the flux values for H- and J-100 assumes that the log activity distance plots for the reactor operating at 100 kw and 1 Mw have the same slopes.

III. RESULTS AND DISCUSSION

A. Approach to Intrinsic Conductivity

The conductivity of all the samples decreases as a function of integrated fast flux and tends toward some limiting value as shown in Fig. 1. This value, σ_L, is tabulated in Table I along with the calculated value of σ_i, the intrinsic conductivity at the temperature of measurement. Values of σ_i were obtained from(14)

$$\sigma_i = p_i e(\mu_e + \mu_h) \qquad (1)$$

where p_i is the density of holes in the intrinsic region and μ_e and μ_h are the electron and hole conductivity

*The number in the sample designation refers to the pre-irradiation room temperature resistivity in units of ohm-cm.

†The assumption is made that the integrated flux, and not the flux, is the important parameter.

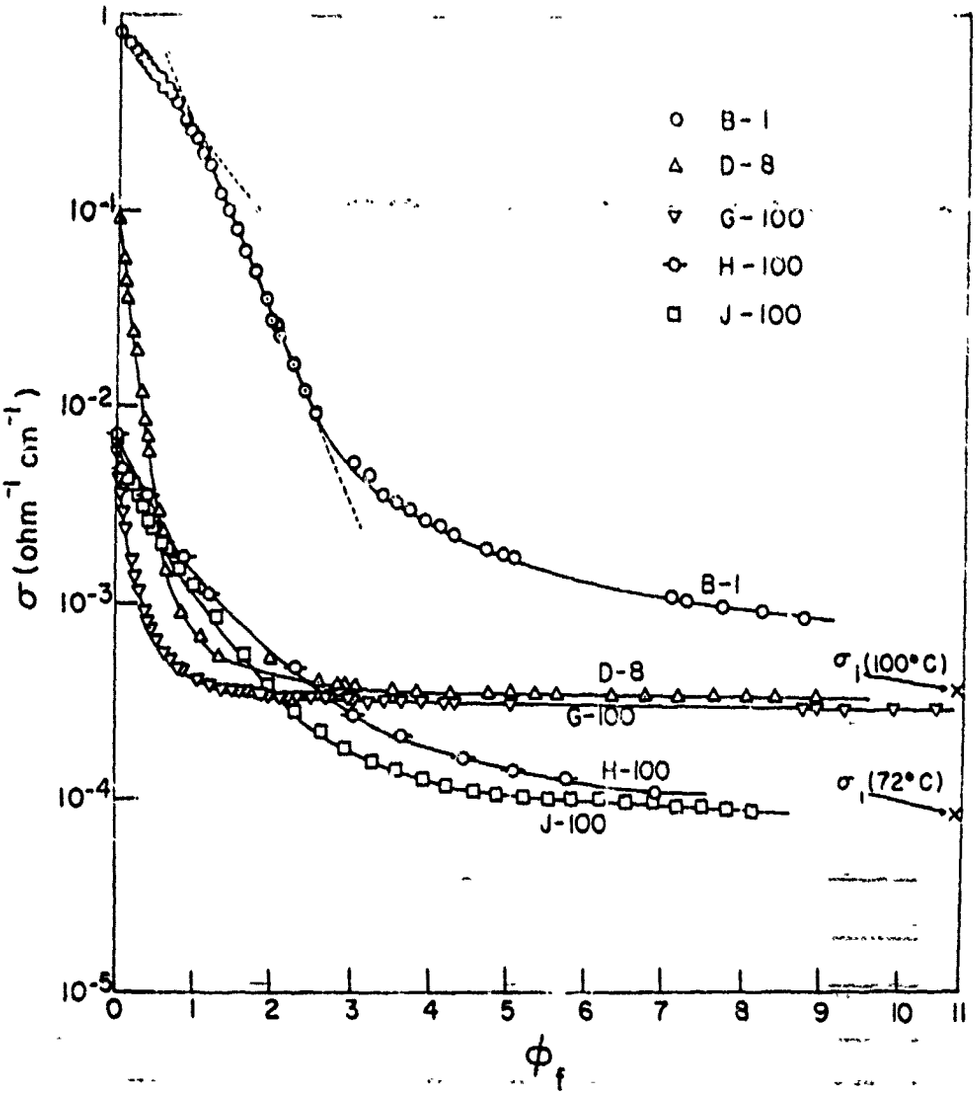


Fig. 1. Conductivity, σ, as a function of integrated fast flux, ϕ_f, for different p-type silicon samples irradiated at 100°C (B-1, D-8, G-100) and at 72°C (J-100, H-100). The crosses are values of σ_i calculated from Eq. (2). To obtain the abscissas of points in n/cm^2, multiply the numbers along the horizontal axis by 2×10^{14} (B-1, D-8), 5×10^{13} (G-100), 2×10^{12} (J-100), and 10^{12} (H-100). Some of the points at very low integrated flux for Sample B-1 have been omitted for the sake of clarity.

mobilities, respectively. If one uses the empirical expressions for the conductivity mobilities and for p_i as obtained by Morin and Maita(15), one obtains

$$\sigma = \left(1.5 \times 10^{33} T^3 e^{-\frac{1.21}{kT}}\right)^{\frac{1}{2}} e\left[1450\left(\frac{300}{T}\right)^{2.6} + 500\left(\frac{300}{T}\right)^{2.3}\right] \quad (2)$$

The temperature range over which Eq.(2) is valid is that for which lattice scattering is the main scattering mechanism. According to Conwell(16), the impurity contribution to the mobilities of the samples used (10^{16}, 10^{15}, 10^{14} impurity atoms/cm^3) is small. The influence of radiation-induced defects on the scattering may be estimated as follows. If one takes(17) 10 displaced atoms/cm^3 per neutron/cm^2*, the total density of displaced atoms, N_d, at the end of irradiation is found and the values are given in Table I. The largest value of N_d is 1.8×10^{16}/cm^3, which is about the same as the initial chemical impurity concentration for B-1. Since an initial impurity concentration of 10^{16}/cm^3 does not have much effect on the mobility, neither should the radiation-induced defects, and the use of Eq.(2) seems quite reasonable. This is more a plausibility argument than a proof since one should really measure μ_e and μ_h as a function of irradiation.

It is clear from Table I that Samples D-8 and G-100 have become intrinsic, H-100 and J-100 are almost intrinsic, and Sample B-1 is not intrinsic. It is to be noted that the sample becomes intrinsic only when the density of displaced atoms is 10 times the initial hole density, p_o, also given in Table I. The significance, if any, of this factor of 10 is not known at present. The purer the sample, the smaller is the initial carrier density, and the smaller is the integrated fast-flux necessary to make the sample go intrinsic. (Compare D-8 with G-100 both irradiated at 100°C.) Also, the lower the temperature, the smaller is the integrated flux needed to make the sample become intrinsic. (Compare G-100 with H-100 or J-100.)

* A value of 37 cm^{-1} was calculated for silicon using the method of reference(17). This value was arbitrarily reduced by a factor of 4 on the basis of a statement also in reference(17), p. 36, to the effect that the ratio of calculated to observed numbers of defects "seems to lie in the range of 3 to 10 for most materials."

TABLE I

Sample Characteristics and Irradiation Data

	B-1	D-8	G-100	H-100	J-100
ρ_o	1	8	100	100	100
p_o	1.9×10^{16}	1.7×10^{15}	1.4×10^{14}	1.3×10^{14}	1.2×10^{14}
σ_L	8.0×10^{-4}	3.2×10^{-4}	2.8×10^{-4}	1.0×10^{-4}	8.5×10^{-5}
σ_i	3.5×10^{-4}	3.5×10^{-4}	3.5×10^{-4}	8.1×10^{-5}	8.1×10^{-5}
E_f-E_v	0.21	0.27	0.37	0.34	0.34
N_d	1.7×10^{16}	1.8×10^{16}	5.2×10^{15}	2.9×10^{14}	1.6×10^{14}
ϕ_{f_1}	---	1.0×10^{14}	2.6×10^{13}	4.5×10^{12}	6.4×10^{12}
α	$\begin{cases}80\\40\end{cases}$	60	20	115	140
P_h(E=0.16ev)	0.17	0.026	0.0014	0.0023	0.0023
P_h(E=0.30ev)	0.94	0.72	0.10	0.21	0.21
ϕ	3.3×10^{10}	3.3×10^{10}	1.1×10^{10}	1.1×10^{9}	3.6×10^{8}
t	100	79-100	100	72	72

ρ_o = nominal room temperature resistivity in ohm-cm; p_o = initial number of holes/cm^3; σ_L = limiting conductivity near end of irradiation in (ohm-cm)$^{-1}$; σ_i = intrinsic conductivity in (ohm-cm)$^{-1}$ calculated from Eq. (2); E_f-E_v = initial Fermi energy, E_f, minus the energy of the top of the valence band, E_v, in ev calculated from Eq. (3); N_d = calculated total number of displaced atoms/cm^3; ϕ_{f_1} = integrated flux at which sample conductivity is 2 σ_i in neutrons/cm^2; α = removal rate in cm^{-1}; $P_h(E)$ = calculated probability of occupancy of levels at E=0.16 ev and 0.30 ev by holes; Φ = flux in neutrons/cm^{2}sec; t = irradiation temperature in °C.

To compare the data of this paper with that of other workers, it is convenient to use the value of the integrated flux, ϕ_{f_1} at which σ equals 2 σ_i; those values are tabulated in Table I. Crawford and Cleland(18) observed the behavior of approach to intrinsic nature of pile-irradiated n- and p-type pulled silicon having initial resistivities of 0.8 ohm-cm and 5 ohm-cm, respectively. Their p-type silicon irradiated at 60°C was within a factor of 2 of being intrinsic after 1×10^{17} n/cm^2. This is a factor of 10^3 greater integrated flux than the value for the 8 ohm-cm sample irradiated at a slightly higher temperature in the present experiment. Longo(19) irradiated 2 ohm-cm n-type and 12 ohm-cm p-type, pulled silicon in a reactor and found the samples to have resistivities corresponding to intrinsic material following the irradiation at a temperature of 45°C. His quoted integrated flux after irradiation was 7×10^{17}n/cm^2. Thus, with respect to the approach to intrinsic conductivity, zone-refined and pulled silicon crystals of equivalent resistivity behave in the same manner except that zone-refined silicon approaches intrinsic conductivity with much less irradiation than pulled silicon. This agrees with a footnote statement of Watkins, et al.(20), who made measurements of electron-irradiated silicon.

B. Detailed Variation of Conductivity with Integrated Fast Flux

1. Sample B-1

Figure 1 shows that the initial behavior of B-1 is different from that of D-8 and the 100-ohm-cm samples. For Sample B-1, there are initially two linear regions of ln σ vs ϕ_f. (To make certain that a similar behavior was not missed in the two 100 ohm-cm samples because of the rapid change of σ with ϕ_f, a third 100 ohm-cm sample, J-100, was irradiated at a very low flux of 3.6×10^8n/cm^{2}sec. The conductivity data shown for this sample in Fig. 1 was taken over a period of 780 minutes and a monotonic non-linear decrease of ln σ with ϕ_f was observed.)

To help explain the behavior of the conductivity as a function of integrated fast flux for Sample B-1 (and for the other samples as well), we will examine the variation of the Fermi energy with integrated flux. Analysis of the Hall data reveals for Sample B-1 that ln p vs ϕ_f, where p is the hole density, also has two linear regions. Furthermore, the hole density is related to the Fermi energy, E_f, by use of(21)

$$E_f - E_v = kT \ln \left[2\left(\frac{2\pi m_h kT}{h^2}\right)^{\frac{3}{2}} \frac{1}{p}\right] \qquad (3)$$

where m_h is the density-of-states mass for the holes in the valence band and E_v is the energy of the top of the valence band. Therefore, since ln p vs ϕ_f is linear, E_f vs ϕ_f will also be linear. In addition, we assume that the lower half of the forbidden band has defect levels similar to those in the neutron-irradiated, pulled silicon observed by Klein and Straub(6). That is, there are levels at 0.16 ev and 0.30 ev above the valence band.* The density of holes trapped by the defects at a given temperature depends on the probability of occupancy of the level and the net rate at which the defects are formed. The probability of occupancy, shown in Table I, is calculated from $[1+\exp(E_f-E)/kT]^{-1}$, and is 0.17 and 0.94 for 0.16 ev and 0.30 ev levels, respectively, for Sample B-1. Thus, if both levels were formed at the same rate, the 0.30 ev level would be populated much faster with holes than the 0.16 ev level, because its energy is lower (as far as holes are concerned) and consequently its probability of occupancy is higher. But the 0.30 ev level is formed at a lower rate (as will be discussed in the following section) than the 0.16 ev level. In the case of the 1 ohm-cm sample, then, the Fermi level increases linearly with ϕ_f until the combination of the two factors above favors the filling of the 0.30 ev levels instead of the 0.16 ev levels. At this point, E_f begins to change with ϕ_f at a different rate and the second linear region of ln σ vs ϕ_f begins. Throughout the entire irradiation, of course, E_f is increasing toward the center of the gap. Finally, as saturation occurs, there is established an equilibrium which involves the density of holes left in the valence band and the density of the 0.30 ev levels being formed. Longo(19) observed more than one slope in ln σ vs integrated flux for 0.05 ohm-cm initial-resistivity (presumably pulled) p-type silicon irradiated with deuterons. The end of the first linear portion of ln σ vs integrated deuteron flux in Longo's experiment occurred when 6 x 10^{17} holes/cm^3 had been removed from the valence band. This would occur when approximately 10^{18} levels/cm^3

*The following discussion in the text does not depend on there being a hole trap at 0.06 ev[8] above the valence band since the probability of occupancy of this level is 0.01 for B-1, 0.001 for D-8, and is much lower for the 100 ohm-cm samples.

which involve oxygen atoms* were filled. In the present work with zone-refined silicon (oxygen concentration less than $10^{16}/cm^3$), the end of the first linear portion occurs when approximately 1×10^{16} holes/cm^3 are removed from the valence band. Thus, the density of holes removed by the time the second linear portion of ln σ vs ϕ_f begins in both pulled and zone-refined silicon seems to be very nearly the same as the available oxygen density.†

2. Sample D-8

For Sample D-8, the Fermi level is initially at 0.27 ev (as claculated from Eq.(3) and shown in Table I) above the valence band, and, also from Table I, one sees that the probability of occupancy of the 0.16 ev and 0.30 ev levels is 0.026 and 0.72, respectively. The probability of occupancy of the 0.30 ev level is therefore 30 times higher than that of the 0.16 ev level, and the difference in production rate of the 0.16 ev and 0.30 ev levels is not sufficient to counteract the much higher probability of occupancy of the 0.30 ev levels. Thus, the single slope in the plot of ln p vs ϕ_f or of E_f vs ϕ_f for

---

*The lower lying 0.16 ev level could result from a hole trap due to various impurities (including boron) other than oxygen. However, since the conductivity of zone-refined silicon changes much more rapidly with irradiation than that of pulled silicon of equivalent resistivity, it seems reasonable that oxygen is the responsible agent, simply because this is the primary difference between zone-refined and pulled silicon. In addition, oxygen was chosen as the atom responsible for the trap at 0.16 ev because of the similarity between the defect levels in the upper and lower halves of the energy gap[8] and because the level 0.16 ev below the conduction band in electron-irradiated pulled n-type silicon is due to an oxygen atom-vacancy combination(11). There is some controversy as to whether neutron-irradiated n-type silicon has any discrete trapping levels. According to Rupprecht and Klein(9) there is a discrete level at 0.15 ev below the conduction band but Wertheim(5) and Sonder(22) claim there is a continuous set of levels in the upper half of the band.

†The density of holes removed during the first linear portion is also the same as the initial boron impurity concentration, and the assumed hole trapping level at 0.16 ev above the valence band could be due as much to a boron-involved level as to an oxygen-involved level. See the footnote above(*) as to why oxygen was chosen as the impurity causing the 0.16 ev level.

D-8 is due to a filling of the 0.30 ev levels. As saturation takes place, the change of ln p with ϕ_f is governed by the rate at which equilibrium is established between the density of the 0.30 ev level defects being produced and the density of holes in the valence band.

3. Samples G-, H-, J-100

In the case of Samples G-, H-, and J-100, the Fermi level is initially 0.34 ev or more above the valence band and the probability of occupancy of the 0.16 ev defect level is 0.0014 for G-100 and 0.0023 for H- and J-100. One can note that the probability of occupancy of the 0.30 ev level is initially 0.10 or higher and, although low, is close to a factor of 100 times that of the 0.16 ev level. Thus for these samples, as for Sample D-8, the larger occupation probability of the 0.30 ev level favors its being filled in preference to the 0.16 ev level. The fact that one does not initially get a linear decrease in the curve of E_f vs ϕ_f or in ln p vs ϕ_f for these samples may be due to a competition for the holes by traps formed at 0.30 ev and by traps at higher energies.

If this interpretation is correct, one should find no occupied 0.16 ev levels in the 1 ohm-cm sample after the end of the first linear portion of E_f vs ϕ_f, and no occupied 0.16 ev levels at any irradiation dosage in the 8 and 100 ohm-cm samples. Two recent papers by Vavilov and Plotnikov(23) and Plotnikov, et al.(24) partially confirm this "prediction". They made photoconductivity measurements at 100°K and concluded that there are levels in 100 ohm-cm, zone-refined, pile-irradiated (10^{13}n/cm^2 with average energy of 1 mev) p-type silicon at 0.30, 0.38, and 0.45 ev above the valence band, and no levels at 0.16 ev. To further confirm these ideas, a simple experiment could be done to locate the hole trapping levels in 1 ohm-cm zone-refined silicon after a short enough neutron irradiation such that ln σ is still in the first linear region.

The fact that the hole density and conductivity decreases immediately with irradiation in the case of the 100 ohm-cm samples may indicate that the deeper lying levels do not result from clusters of vacancies or interstitials. Or, if the levels are caused by clusters, the clusters are immediately formed when irradiation is begun.

C. Hall Coefficient Variation with Integrated Fast Flux; Hole Removal Rates

The variation of the Hall coefficient with integrated fast flux, as shown in Fig. 2, can, of course, be explained on the basis of carrier removal and by resorting to the two carrier model as the material becomes intrinsic.

The most interesting feature of the Hall coefficient data is that R, initially positive, reaches a maximum as a function of integrated flux, and then goes negative for Samples G-, H-, and J-100. Similar results were observed with pulled silicon by Longo(19) after much higher irradiation (10^{17}n/cm^2). With deuteron irradiation of 0.16 ohm-cm p-type (presumably pulled) silicon, Longo also observed R to go negative after 7 x 10^{14} deuterons/cm^2 (9.6 mev) at room temperature. Klein and Straub(6), on the other hand, irradiated 60 ohm-cm (presumably pulled) p-type silicon in the Brookhaven reactor at dry ice temperature and observed the Hall coefficient and conductivity after room temperature storage extending from several days to a year. They saw no change in the sign of R with an integrated flux of 10^{14}n/cm^2. In the present work R becomes negative after 1.7 x 10^{13} and 1.4 x 10^{13}n/cm^2 for Samples H-100 and J-100, respectively, and after 7.2 x 10^{13}n/cm^2 for sample G-100. These considerations seem to suggest that R goes through a maximum and becomes negative more rapidly for floating zone-refined silicon than for pulled silicon of equivalent resistivity.

Samples H-100 and J-100, irradiated at different flux rates, show similar dependence of R on integrated flux as seen in Fig. 2. Sample J-100 has a smaller maximum than H-100, probably because of a small initial increase of 10°C in temperature of the former. After a few minutes of irradiation, the temperature of J-100 was back to 72°C. However, the positions of the maxima of R occurred at the same value of ϕ_f and the positions of R = 0 occurred at approximately the same value of ϕ_f for both samples. Sample G-100 was irradiated at a different flux as well as at a higher temperature (100°C) than either H-100 or J-100. The maximum in R for G-100 occurred at an integrated flux of 4 times that for Samples H-100 and J-100. In addition to this, R for G-100 did not become as negative as that of H-100 and J-100 with the integrated flux obtained during the experiment. The small difference in irradiation temperature (on an absolute scale) of 30°C between G-100 and H-100 or J-100 apparently made a large change in annealing rate and hence in the net

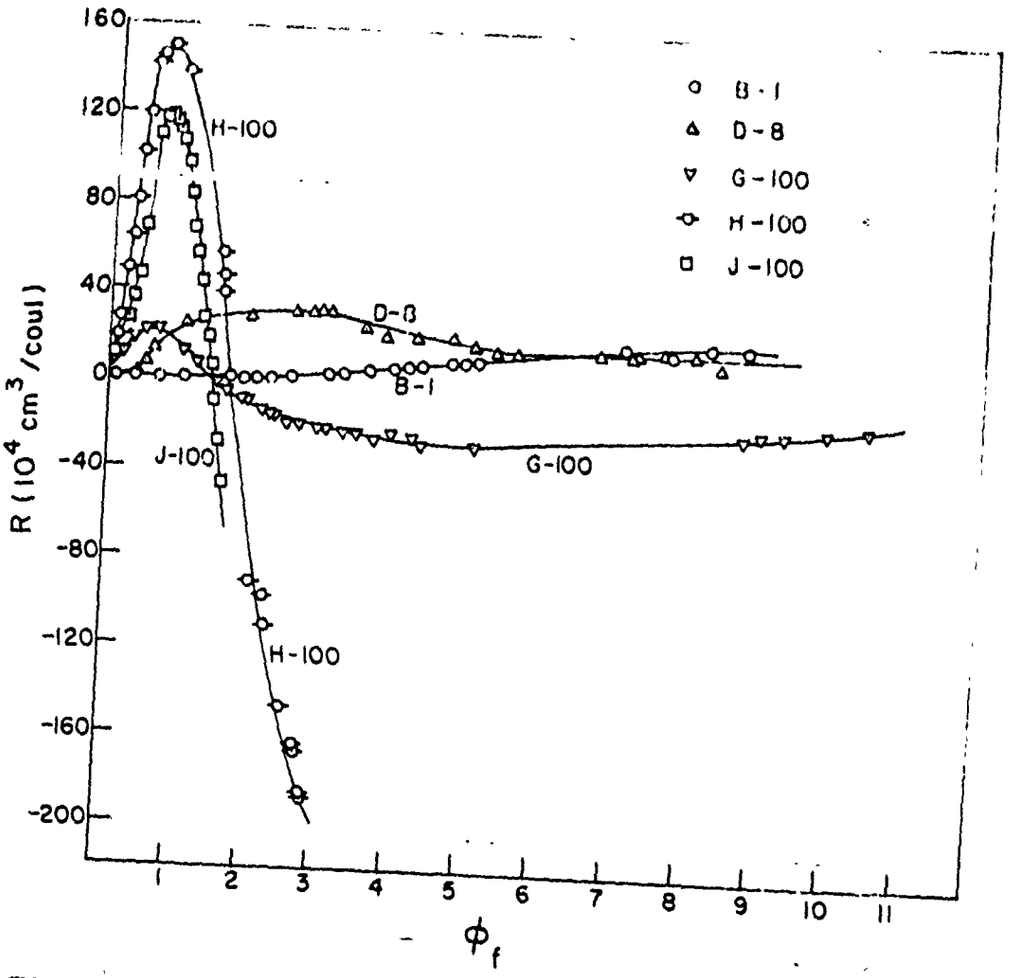


Fig. 2. Hall coefficient, R, as a function of integrated fast flux, ϕ_f, for different p-type silicon samples irradiated at 100°C (B-1, D-8, G-100), and at 72°C (J-100, H-100). To obtain the abscissas of points in n/cm^2, multiply the numbers along the horizontal axis by 2×10^{14} (B-1, D-8), 5×10^{13} (G-100), and 10^{14} (H-100, J-100). Some of the points at very low integrated flux have been omitted for the sake of clarity.

rate of formation of the defect levels. This change was sufficient to increase the required flux necessary to bring the Hall coefficient to zero for Sample G-100.

Sample D-8 exhibits a much broader peak in R vs ϕ_f and remains positive during the entire irradiation time used for this sample. In addition, the other low-resistivity sample, B-1, never exhibited a maximum in R. Longo(19) observed R of 0.16 ohm-cm material to go negative after a long deuteron irradiation. This would suggest that R for the 1 ohm-cm and 8 ohm-cm samples eventually would have gone negative if longer irradiations had been used.

To examine the carrier removal rates we consider the Hall coefficient from which the carrier density for p-type silicon is calculated using $R = r/pe$, where r is the ratio of μ_H, the Hall mobility, to μ_C, the conductivity mobility. This expression can be used only during the initial period of irradiation and is not valid when the sample is approaching intrinsic behavior. According to the graph of Morin and Maita(15) for r vs T, r is about 0.71 and 0.68 for T = 72°C and 100°C, respectively. Because of the spread in their data, a value of 0.70, independent of irradiation, has been taken to compute p, the hole density. The carrier removal rates are listed in Table I. The values are obtained from the first few points of a linear plot of p vs ϕ_f which is consistent with the errors involved. These values are, of course, very approximate because of the uncertainty in the integrated fast neutron flux. The relative values, however, are significant. The initial removal rate for B-1, 80 cm^{-1}, is due primarily to holes being trapped by the 0.16 ev level. At the end of the first linear portion of ln σ vs ϕ_f, the removal rate for Sample B-1 has dropped to 40 cm^{-1} because the holes are now preferentially trapped by defects with energy levels at 0.30 ev. In the case of Sample D-8, the initial removal rate of 60 cm^{-1} is lower than the initial removal rate of B-1. The fact that the holes initially are being trapped primarily by defects with energy levels at 0.16 ev for Sample B-1 and at 0.30 ev for Sample D-8 suggests that the 0.30 ev level defects are being formed at a lower rate than the 0.16 ev levels. The difference between the initial removal rate of D-8 and the removal rate of B-1 at the beginning of the second linear region, when, presumably, for both samples the 0.30 ev levels are being populated, may be due to the lower initial temperature of D-8. Further evidence for the lower rate of formation of the 0.30 ev levels lies in the small initial removal rate, 20 cm^{-1}, for G-100. In this sample, only the defects with energies at 0.30 ev

or higher can trap holes as discussed in the previous section. For Sample G-100, the initial removal rate is lower than for D-8 and for the removal rate at the beginning of the second linear portion of B-1. This may be due to the lower probability of occupancy (0.1) of the 0.30 ev level for G-100 as well as to a competition for holes from the valence band between the 0.30 ev level defects and higher level defects. This seems to indicate that the higher level defects are formed at an even lower rate than the 0.30 ev level.

The large difference in removal rates between Sample G-100 (20 cm^{-1}) irradiated at 100°C and Samples H-100 (115 cm^{-1}) and J-100 (140 cm-1) irradiated at 72°C indicates either a large amount of annealing or perhaps a reduction in the clustering of defects which may lower their effectiveness in trapping. Additional work has to be done to clarify this point.

D. Hall Mobility

The Hall mobility, μ_H, is shown in Fig. 3 as a function of integrated fast flux. The dashed line in Fig. 3 connects the 100°C pre-irradiation value of μ_H for Sample D-8 with the value observed after the sample reached 100°C during the irradiation.

To explain the dependence of μ_H on ϕ_f, we consider the two carrier expression for the Hall mobility:

$$\mu_H = R\sigma = \mu_h \, r \, \frac{p - n\mu_e^2/\mu_h^2}{p + n\mu_e/\mu_h} \qquad (4)$$

As irradiation proceeds, the hole traps cause a decrease in p. As p decreases and approaches $n\mu_e^2/\mu_h^2$, μ_H decreases, goes through zero, and becomes negative. If the 1 ohm-cm sample had been irradiated long enough to approach intrinsic conductivity with its Hall coefficient becoming negative (both conditions were true for the 8 ohm-cm and 100 ohm-cm samples), then μ_H for it would have decreased with irradiation.

SUMMARY

The variations in the Hall coefficient and the conductivity of zone-refined silicon undergoing neutron irradiation are similar to those of pulled silicon except that the changes with integrated flux occur much more rapidly for the zone-refined material. The log of the

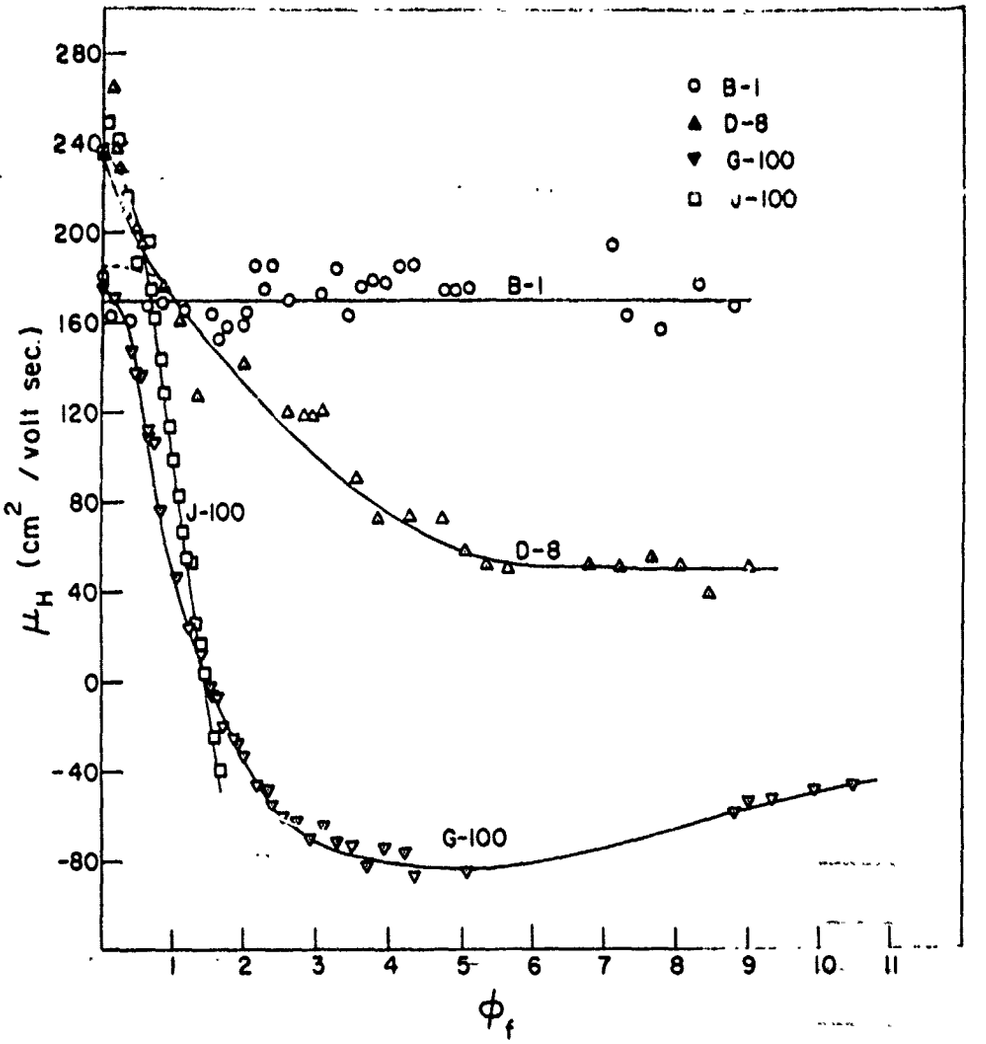


Fig. 3. Hall mobility, μ_H, as a function of integrated fast flux, ϕ_f, for different p-type silicon samples irradiated at 100°C (B-1, D-8, G-100) and at 72°C (J-100). To obtain the abscissas of points in n/cm^2, multiply the numbers along the horizontal axis by 2×10^{14} (B-1, D-8), 5×10^{13} (G-100), and 10^{13} (J-100). Some of the points at very low integrated flux have been omitted for the sake of clarity.

conductivity shows a monotonic, non-linear decrease with irradiation for the 100 ohm-cm samples, a single region of linear decrease for the 8 ohm-cm sample, and two regions of linear decrease for the 1 ohm-cm sample. For the latter sample, this effect is attributed partly to the filling of the levels which may involve oxygen atoms in some way. The initial hole removal rates are 80, 60, 20 cm^{-1} for the 1, 8 and 100 ohm-cm samples, respectively, at an irradiation temperature of 100°C, and 115 cm^{-1} for one of the 100 ohm-cm samples at an irradiation temperature of 72°C. The Hall coefficient goes through a maximum and decreases for the 8 and 100 ohm-cm samples and then goes through zero and becomes negative for the 100 ohm-cm samples. In the case of the 1 ohm-cm sample, the Hall coefficient increases and then levels off. The Hall mobility decrease in the 8 and 100 ohm-cm samples is ascribed to a decrease in the carrier density. The Hall mobility is constant with irradiation for the 1 ohm-cm sample.

ACKNOWLEDGMENT

The authors are deeply indebted to Dr. A. I. Schindler who helped clarify many points by invaluable discussions and his assistance in reading and correcting the manuscript. We are also grateful to Dr. E. I. Salkovitz for his guidance and support during the initial stages of this work. Several discussions with Dr. C. A. Mackliet were helpful. Finally, we want to thank the NRL reactor group headed by Dr. J. O. Elliot for their cooperation during the in-pile measurements.

REFERENCES

1. G.K. Wertheim, Phys. Rev. 105, 1730 (1957).

2. G.K. Wertheim, Phys. Rev. 110, 1272 (1958).

3. D.E. Hill and K. Lark-Horovitz, Bull. Am. Phys. Soc., Ser. II, 3, 142 (1958).

4. T.A. Longo and K. Lark-Horovitz, Bull. Am. Phys. Soc., Ser. II, 2, 157 (1957).

5. G.K. Wertheim, Phys. Rev. 111, 1500 (1958).

6. C.A. Klein and W.D. Straub, "Proceedings of the Third Semi-Annual Radiation Effects Symposium," Oct. 28, 29 and 30, 1958, Sponsored by ARDC, USAF. Lockheed Nuclear Products, Paper No. 30.

7. D.E. Hill, Phys. Rev. 114, 1414 (1959).

8. Claude A. Klein, J. App. Phys. 30, 1222 (1959).

9. G. Rupprecht and C.A. Klein, Phys. Rev. 116, 342 (1959).

10. R.K. Willardson, J. App. Phys. 30, 1158 (1959).

11. G.D. Watkins and J.W. Corbett, Phys. Rev. 121, 1011 (1961).

12. W. Kaiser, P.H. Keck, and C.F. Lange, Phys. Rev. 101, 1264 (1956).

13. G.C. Bailey and C.M. Williams, Rev. Sci. Instr. 32, 1412 (1961).

14. See, for example, Charles Kittel, Introduction to Solid State Physics (John Wiley and Sons, Inc., New York, 2nd Edition, 1957), p. 351.

15. F.J. Morin and J.P. Maita, Phys. Rev. 96, 28 (1954).

16. Esther M. Conwell, Proc. Inst. Radio Engrs. 40, 1327 (1952).

17. Douglas S. Billington and James H. Crawford, Jr., Radiation Damage in Solids (Princeton University Press, N. J., 1961), p. 34.

18. Unpublished data of J.W. Cleland, J.H. Crawford, and J.C. Pigg quoted in J.H. Crawford and J.W. Cleland, Progress in Semiconductors, Edited by A.F. Gibson, P. Aigrain, and R.E. Burgess (Heywood and Co., London, 1959), Vol. 2, p. 67.

19. T.A. Longo, Ph.D. Thesis, Purdue University, 1957 (Unpublished).

20. G.D. Watkins, J.W. Corbett, and R.M. Walker, J. App. Phys. 30, 1198 (1959).

21. R.A. Smith, Semiconductors (Cambridge University Press, London, 1959), p. 82.

22. E. Sondor, J. App. Phys. 30, 1186 (1959).

23. V.S. Vavilov and A.F. Plotnikov, Soviet Phys. - Solid State 3, 1783 (1962).

24. A.F. Plotnikov, V.S. Vavilov, and L.S. Smirnov, Soviet Phys. - Solid State 3, 2363 (1962).

SOURCES OF STRENGTH AND WEAKNESS IN GLASS REINFORCED PLASTIC STRUCTURES

G. R. Irwin and J. A. Kies
U. S. Naval Research Laboratory
Washington 25, D. C.

Typically the initial stages of failure of a composite structural material occur by cracking at boundaries between essentially different constituents of the composite. Although this is not a scholarly definition of a composite, it serves that purpose in a practical sense for a discussion where mechanical strength is of primary interest. Composite materials are increasing in importance particularly where a high ratio of strength to weight is needed. In some composites cracking at weak boundaries is a sufficiently general statement of the failure mechanism. In other composites, notably in glass reinforced plastic structures, the failure processes are complex and a detailed understanding of the principle failure mechanisms is needed as a basis for reliability control procedures.

Concrete is a good example of a composite for which the failure mechanism is simple. The applications of fracture mechanics to concrete by Kaplan (1) and Glucklick (2) were assisted by suggestions from the authors of this paper. Portions of the concrete which are large relative to the nonuniformities possess a limited average crack toughness $\mathcal{G}_{Ic}$, of about the same magnitude one finds for window glass. The bonding of large smooth pebbles to the cement mortar is relatively weak and tensile failure occurs when one largest boundary crack becomes of critical size. A similar process occurs in compression failure but the applicable $\mathcal{G}_c$ value, primarily $\mathcal{G}_{IIc}$, is much larger because of interlocking on the plane of shear failure. The tensile failure mechanism occurs in concrete with sufficient regularity so that one can estimate the critical $\mathcal{G}$ value in a test of an unnotched bar simply from the failure load and the estimate of crack size provided by the change in compliance of the specimen at the point of failure.

This has been done by Glucklick (2) in simple beam tests. As the load was applied to the unnotched beam, a linear stress-strain or load-deflection curve was generated which changed slope as the unseen crack or network of cracks developed. Typical test results are shown in Fig. 1. Slow crack growth was occurring at driving forces $\mathcal{G}$ less than $\mathcal{G}_c$; at sudden fracture the "crack depth c" could be estimated from a previous calibration of compliance vs notch depths made within the linear portions of load deflection curves. Glucklick used the Griffith formula

$$\mathcal{G}_c = \frac{\pi c_s \sigma_{st}^2}{E} \qquad (1)$$

and found $\mathcal{G}_c = 0.1100$ lb/in. where $\mathcal{G}_c$ is the Irwin-Griffith critical driving force and σ_{st} is the stress normal to the crack at instability. E is the Young's modulus for the composite with ostensibly no cracks present. The remarkable thing about this was that $\mathcal{G}_c$ determined thus agreed very well with that obtained using deep notches of directly measured depth.

A similar dependence of strength on apparent modulus was found by J. H. Brown (3) in the case of sea ice. Brown did not carry out the calibration experiment but qualitatively his results fit the theory that strength is related to stiffness by the equation (4):

$$\mathcal{G}_c = \frac{P^2}{2} \frac{d\left(\frac{1}{M}\right)}{dA}, \qquad (2)$$

where P is the load, 1/M is the compliance and dA is the increment of crack area associated with d (1/M).

A glass fiber wound chamber subjected to internal pressure may well represent an opposite extreme of complexity. Consider first mechanical failure of a rocket chamber under internal pressure. One may note that there are a certain number of large adhesive joints between the windings and the metal inserts needed for attachment of igniter and nozzles. Trouble with these large adhesive joints can be minimized by introducing allowable amounts of flexibility and avoiding unnecessary stiffness.

Typical failures due to cracking in the epoxy bond between aft cluster plate are shown for Polaris second stage chambers HFG Nos. 5 and 6 in Figs. 2 and 3. Such failures were marked by load bangs when the bond started to crack well before final rupture. The use of an elastomeric bonding has eliminated this kind of failure. A more gradual load transfer is thereby achieved at the boundary of the joint. A considerable advance in the theory and practice of how to do this is given in Ref. (5). More research is needed to optimize such joints. Fracture mechanics of bonded joints is being developed at NRL and on contract under the direction of G. R. Irwin. The energy for debonding has been shown to be a function of joint thickness and

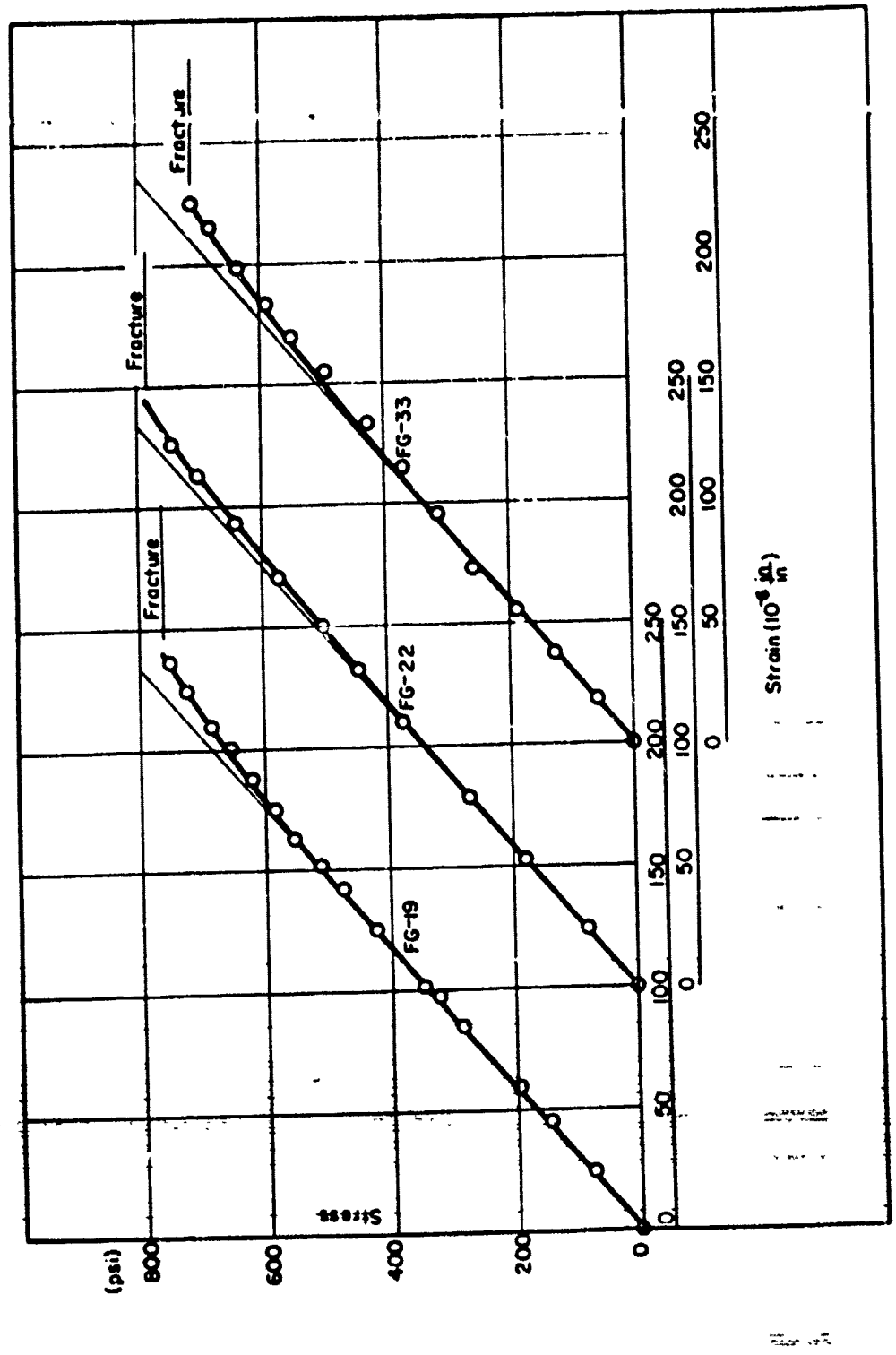


Fig. 1 Static stress strain curves for unnotched beams of concrete. Note that the slope of the stress strain curve at the fracture point could be used to calculate the strength.

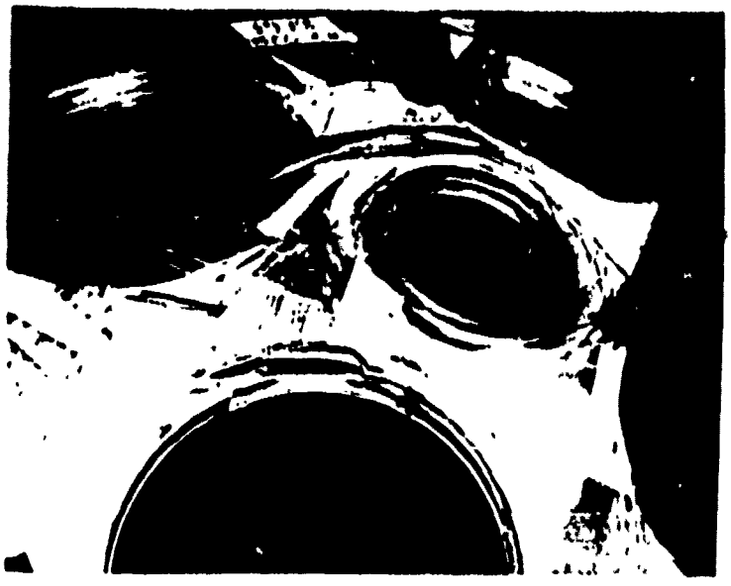


Fig. 2 Aft head failure in epoxy bond between cluster plate and g.r.p. B.F.G. No. 5 Polaris 1nd stage A2



Fig. 3 Aft head failure in epoxy bond between cluster plate and g.r.p. B.F.G. No. 6 Polaris 2nd stage A2

of whether the fracture path is in the cement or at the interface between metal and cement (6). Fracture mechanics offers a new and powerful method for use in evaluating and guiding the development of better bonded joints.

Figure 4 illustrates the kind of specimen recommended for research on bonded joints. $\mathcal{G}_I$ for direct opening and $\mathcal{G}_{II}$ and $\mathcal{G}_{III}$ for shear were defined in Ref. (7). In accordance with Fig. 4, the forces P are exerted through strong fingers separated by a stiff mechanical arrangement virtually equivalent to a wedge. The displacement separation Y of the forces P relative to the zero load position was measured for various crack lengths a as a function of the applied load. Thus, the compliance C (per unit specimen thickness) could be calculated in accordance with the equation

$$Y = C\,P/B \tag{3}$$

where B is specimen thickness. From simple beam theory, one anticipates

$$C = A\,(a + a_o)^3 \tag{4}$$

where

$$A = 8/EBD^3 \tag{5}$$

D = depth of beam
a_o = radius of loading hole
E = Young's modulus of adherent material.

The compliance values found were in essential agreement with Eq. (4) when some allowance was made for an "elastic foundation" contribution to the deflection from the adhesive. The adhesive used was a commercial epoxy often employed as a photo-elastic coating. Comparisons were made with the compliance of a solid aluminum bar with various crack sizes represented by saw cuts. For purposes of this discussion, it may be assumed that compliance data in an experiment of this kind can be represented well enough by

$$C = A\,(a + a_o'')^3 \tag{6}$$

where a_o'' is moderately greater than a_o and represents the effect upon compliance of the strains in the low modulus component, material B.

The crack extension force is given by

$$\mathcal{G} = \frac{1}{2}\,\frac{P}{B}^2\,\frac{dC}{da} \tag{7}$$

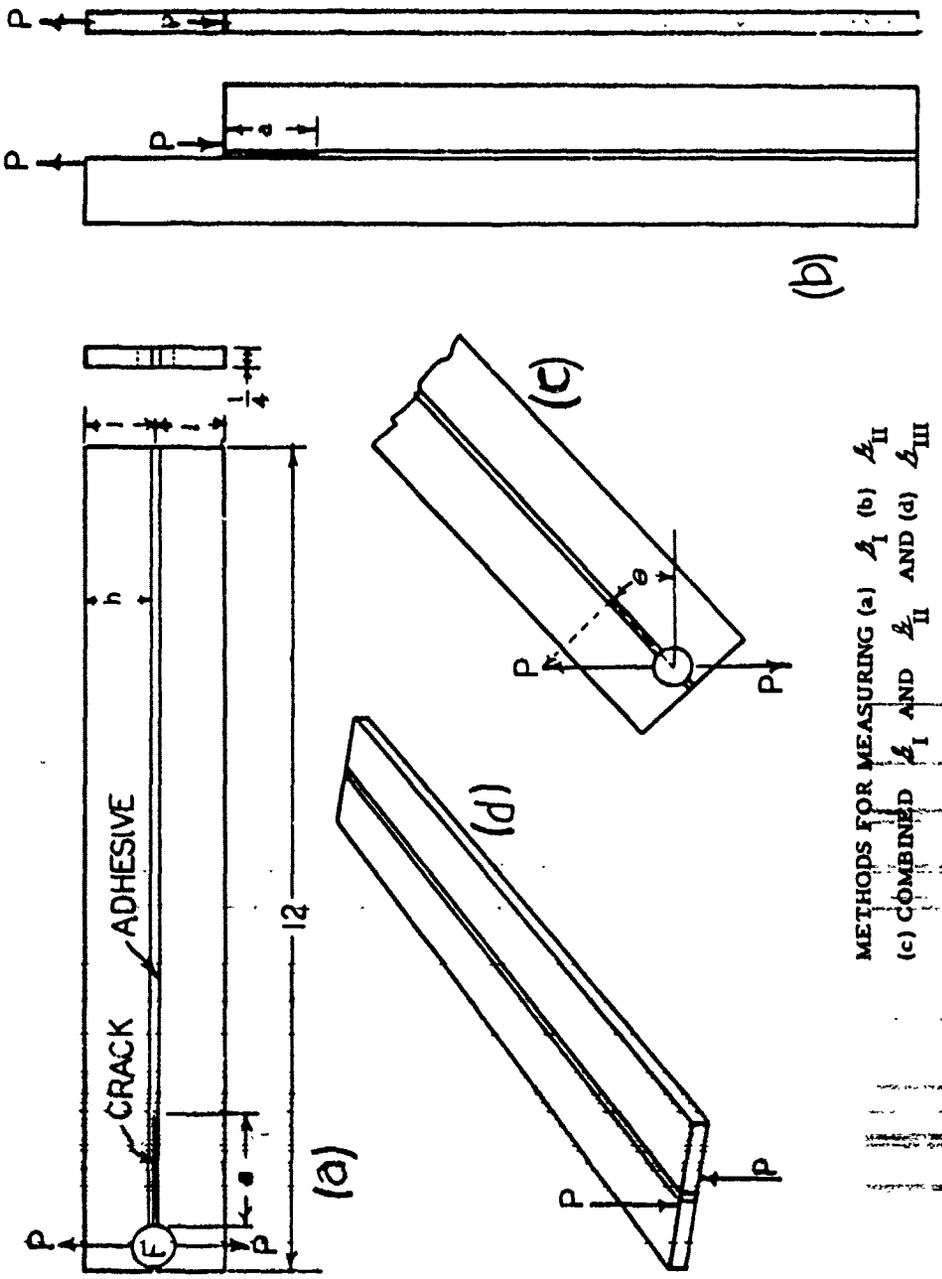


Fig. 4 Specimens and manner of loading of specimens for the fracture mechanics of bonded joints

On the basis that the effective crack length during test of an adhesive joint is the value of a consistent with Eq. (7) and with the observed values of P and Y, one can compute $\mathcal{G}_I$ for any pair of P, Y observations from the equation

$$\mathcal{G}_I = \frac{3}{2}\frac{PY}{B}\left(\frac{AP}{BY}\right)^{1/3} \qquad (8)$$

With Y fixed, the fractional change of crack-extension force with a is given by

$$\frac{1}{\mathcal{G}}\frac{\partial \mathcal{G}}{\partial a} = -\frac{4}{a + a_o''} \qquad (9)$$

Adhesive joints, as customarily prepared, are nonuniform with regard to such factors as joint thickness, local residual stress, voids, and surface condition. In addition, a high polymer adhesive is normally strain rate sensitive in such a way that it tends to respond with greater stiffness and less toughness to rapid straining such as might occur from a rapid increment of crack extension. In attempting to make specimens and measurements of the type discussed here, one will find considerable care in joint preparation is necessary in order to produce a specimen which can be partially precracked (say by tapping with a sharp knife blade) without complete separation of the joint. Even a carefully prepared joint will require a certain "stability factor". The preceding term will be used for the left side of Eq. (9). For the aluminum-epoxy-aluminum joints studied by Ripling, Patrick, and co-workers, unstable rapid fractures occurred when a negative magnitude of the stability factor was appreciably less than 0.5 reciprocal inches. In consistency with Eq. (9), stable crack extension was observed from the pre-crack position to a crack length of nearly 8 inches. In order to study substantially weaker, less uniform joints, one could employ beams of smaller depth.

Measurements of P and Y were also made with the forces tilted to a 45 degree angle relative to the joint. By using the components of P and Y normal and parallel to the joint, it was possible to estimate the division of the total $\mathcal{G}$ value between $\mathcal{G}_I$ and $\mathcal{G}_{II}$. With shearing strains present parallel to the joint, not only was the total $\mathcal{G}$ much larger but also the value of $\mathcal{G}_I$ was several-fold greater than had been measured under opening mode conditions.

When adhesive joints are used in structural components, it is customary to avoid, if possible, loads other than compression and shear on the adhesive joints. Since this is often possible only to an imperfect degree, the critical $\mathcal{G}_I$ value for an adhesive joint is not entirely of academic interest. However, somewhat greater attention is usually given to the strength relative to shear separation. Several values of $\mathcal{G}_{II}$ were measured with forces applied as indicated

in Fig. 5. For this experiment, the compliance is a linearly increasing function of crack length and $\mathcal{G}_{II}$ may be computed from

$$\mathcal{G}_{II} = \frac{1}{4}\left(\frac{P}{B}\right)^2 \frac{1}{ED} . \qquad (10)$$

The resulting $\mathcal{G}_{II}$ values were an order of magnitude greater than the $\mathcal{G}_I$ values. Although the instability restraint factor was only half as large as for the Mode I tests, a long stable movement of the crack was observed.

It is noteworthy that a crack in a plate of glass is easily extended in a controlled stable fashion by use of splitting forces near the plane of the crack. Although the necessary crack-extension force $\mathcal{G}_I$ is less than 0.1 lbs/in., only a small negative magnitude of the stability factor permits control of the crack speed. It would appear, therefore, that strain rate sensitivity coupled with non-uniformities in the adhesive joint, rather than low crack toughness per se, is the basis for the need of a large negative stability factor in the adhesive joint test specimen.

At the present time measurements of $\mathcal{G}_{Ic}$, $\mathcal{G}_{IIc}$, and the associated stability factors provide at least a comparative measure of the danger of crack propagation in service component adhesive joints. Additional studies are needed to classify the necessary test conditions and to give closer quantitative meaning to the results.

When mechanical failure does not begin with these large adhesive joints there remain numerous failure mechanisms of other kinds. One class of these consists in concentrations of glass fiber breaks which develop in a linear pattern along the edge of a stiffening element, a reinforcement doily, a wrinkle in the insulator lining, or a displaced roving. This is largely a matter of stress concentration due to introduction of local bending. Although flaws of this type can be recognized, it is not clear they can be altogether avoided. The material must possess some tolerance for stress irregularities.

An example of failures due to stress concentration in combined tension and bending is shown in Fig. 6. The Aerojet-General Corp. solution to this problem was straightforward, but consisted of five simultaneously applied remedies: (1) reversing the order of lay up of reinforcing doilies according to size; (2) moving the skirt fillet away from the high bending moment region; (3) filling the aft skirt fillet with an elastomer; (4) smoothing out the trajectories of the longitudinal windings by applying glass resin fillers where needed and; (5) using a higher strength glass S-994. This example is given to illustrate the necessity of immediate empiricism when conditions are too complicated for detailed analysis. A major improvement of 45 percent in burst strength is believed to have been accomplished by these remedies. Although the remedies were somewhat empirical, the

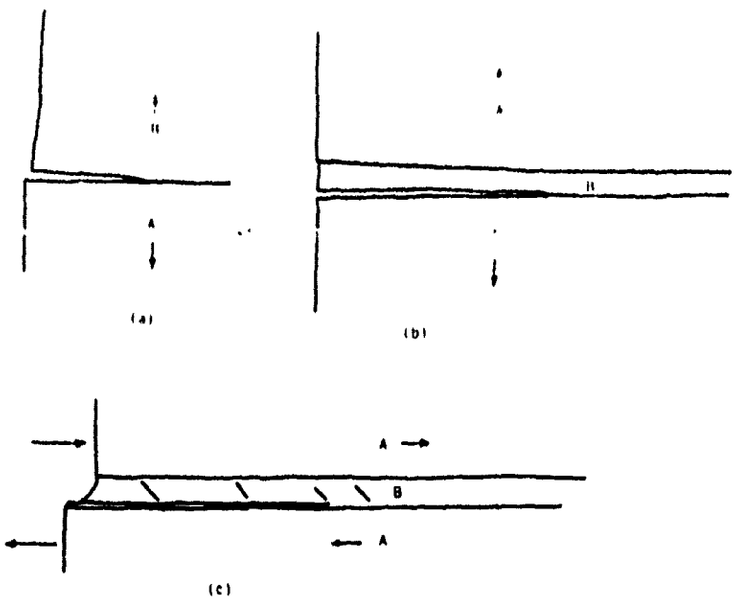


Fig. 5 Separations from the left boundary in adhesive joints



Fig. 6 Aft closure failure in Polaris A3 first stage due to stress concentration at the edge of a stack of doilies. Note that a fairly straight tensile break occurred around the entire periphery. The diameter of the chamber is 54 inches.

improvement of design depended upon thorough examination and analysis of full scale failures.

The development of a separation through a multitude of parallel glass fibers bonded with epoxy resin is, of course, relatively easy along paths parallel to the fibers. On the other hand, if the region is subjected only to uniform tension parallel to the fibers, a tensile break may require a stress nearly as large as the measured tensile strength of single fibers representative of the glass in the region considered. This quality which is basic to the high strength of GRP composites develops in the following way. As a crack in the resin approaches a glass fiber, debonding occurs near the crack plane. This relieves the stress concentration, encourages a fiber fracture away from the plane of the resin crack, and the fracture occurs at a tensile load reflecting the very high strength of a short fiber segment with its correspondingly low flaw probability. At the same time this action causes development of numerous interrupted cracks rather than a single clear break. The effect of filament breaks on the strength of composites has been discussed by Zender and Deaton (8) and by Hedgepeth (9).

A similar toughening tendency occurs to various degrees through the tendency of adjacent rovings to debond from each other as strains become very large and to debond between adjacent layers at boundaries separating fibers wound at a different angle. Although debonding may appear to be a crude scheme for introducing toughness, numerous successful rocket chambers testify this procedure can be of practical value.

On the other hand the "shredded" appearance of typical burst failures of fiber wound rocket chambers suggests the debonding action is somewhat overdone and that introduction of more flexible load transfer couplings would simultaneously reduce the amount of debonding and strengthen the structure. An extreme case of debonding between layers in an aft closure is shown in Fig. 7. This is known as a "flower burst" and is undesirable.

The role of the resin or other matrix material in an optimum composite is still uncertain and sometimes controversial. We need the crack arresting feature but we also need strength in the matrix. Designers of GRP pressure bottles have, at present, no valid way for calculating the required strength of the matrix material. To emphasize the importance of this we now refer to Figs. 8, 9, and 10. The currently used epoxy system is obviously much better than gum rubber or paraffin in which debonding and splitting between fibers is relatively easy even in a simple isotensoid bottle. Low burst strength was obtained using rubber and paraffin as matrices although tensile fracture in the glass was virtually absent.



Fig. 7. "Flower" burst in aft head. This consisted largely of delamination between layers of windings. Much of the thickness in the aft closure is cut filaments in the form of doilies. These can only carry membrane load provided shear stresses can be transmitted across the interface between layers. Interlaminary debonding overloads the longitudinal windings.



Fig. 8 Typical burst in a g.r.p. Outwater test bottle Z-4 using Epoxy 828 resin and E glass. Burst pressure 1157 psig. Note the mixture of tension failures and resin splitting between fibers. The Outwater test bottle (Figs. 8–10) is a natural isotensoid 7.75 inches in diameter at the equator of the oblate ellipsoid. The windings spiral around the pole pieces at 20 degrees from the pole. For the bottles shown the load F per "end" of glass is

$$F = \frac{\pi d^2 p}{8 n N \cos 20^\circ}$$ where

n is the number of ends (here 12) per roving,

N = the number of helical turns at 20° from the pole,

d = the equatorial diameter and p = the gage pressure.



Fig. 9 Burst in an Outwater bottle U-5 using gum rubber as the matrix. Almost no tensile failures in glass occurred. Failure was by debonding between glass and the matrix and by shear displacements separating the fibers. Burst pressure 589 psig. The same amount of glass was used as in Z-4.



Fig. 10 Burst in Outwater test bottle II-8 using paraffin as the matrix. Debonding between glass and matrix was extremely easy and no tensile failures occurred in the glass. Failure was by separations between fibers and shear displacements. Burst pressure 405 psig.

Craze cracking in the present epoxy matrix during proof testing is now very intense. This produces a porous structure and although it has crack arresting merit, the porosity can admit moisture. The cracking is largely the result of strain concentration in the resin matrix during tensile strains transverse to the fibers and in shear. The maximum strain concentration factors in a pressure bottle of balanced construction a very nearly the ratios of Young's moduli for tension or compression and of the shear moduli for shear strains. This situation has been explained in detail in previous reports, (10, 11, and 12). It is clear that strains as high as forty percent would be imposed on the uncracked resin for a two percent overall strain transverse to the fibers. Cracking occurs well before this point is reached. Figure 11 illustrates the external appearance of a large rocket case after proof test and Table I gives typical values of computed strain magnifications in resin.

TABLE I

Effect of Glass Content on Tensile Strain Concentration in the Resin, Square Array, Assuming $E_g/E_r = 20$

Δ/R	Glass Volume Fraction	$\epsilon_{xr}/\bar{\epsilon}_x$
1.0	0.349	2.73
0.5	0.503	4.17
0.2	0.650	7.33
0.1	0.713	10.5
0.05	0.748	13.6
0.02	0.770	16.8
0.01	0.778	18.3
0	0.786	20

Although the glass fiber itself is the basic source of strength, this paper is primarily concerned with failure mechanics and so will not discuss fiber strength in any detail. The increase of strength from use of S-994-HTS fibers in place of E-HTS glass is substantial and still higher glass fiber strengths appear attainable in the future. Nevertheless, for perspective, one can bear in mind that chambers constructed in what was thought to be proper fashion with S-994 glass have occasionally failed in hydrotest. In addition, it has been shown that a substantial increase in the parallelism of the fibers in a glass roving provides a strength increase as large as the fiber strength margin of S-994 over E glass. Lack of parallelism means the failure of crooked or slack fibers to carry their share of the load. An example of this defect is shown in Fig. 12. The extent to which the resin holds the fibers fixed in position partly determines the tolerance of the structure for this defect. Evidently efficient utilization of higher strength fibers requires careful

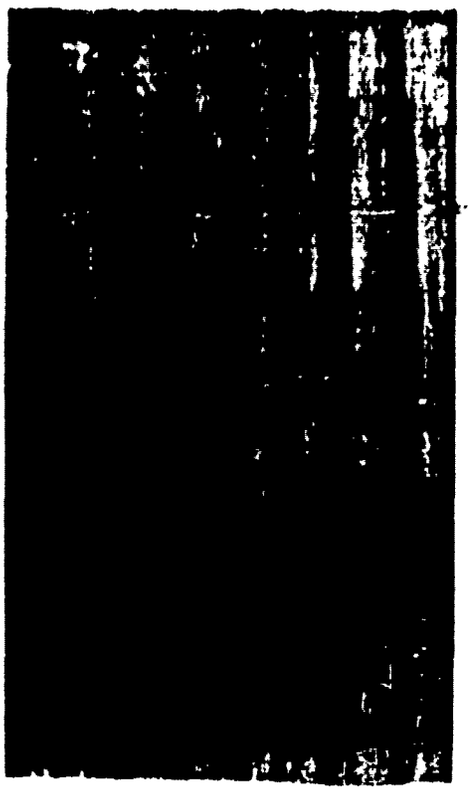


Fig. 11 Craze cracking in the resin after proof test of to 1.5 percent average strain transverse to the windings shown, first stage Polaris



Fig. 12 Lack of parallelism in the windings of a pressure bottle. This sectional view shows crookedness in hoop windings produced by poor tensioning control. This bottle had only one-half the strength of a nearly perfectly wound bottle of the same design. Stronger and stiffer resins tend to compensate for such a defect.

control of the factors which degrade the basic strength possibilities and lead to occasional hydrotest failures.

Mechanical damage which results in the cutting or easy breakage of a considerable number of fibers on the surface can result in early failure. Due to the crack arresting properties related to debonding, a crack does not necessarily propagate immediately across adjacent fibers; however, a progressive failure can result in which shear cracking in the resin is an important step. Figure 13 is an example of such "unwinding" failure in a notched bottle. When no intentional notch is present, very small strands break, usually at more than one position. From these small tensile breaks, unwinding failure develops as shown in Fig. 14. This is one of a series of six photographs obtained showing a gradual spread of the damage during a three minute hold at constant pressure at a little below failure pressure. The photograph supplied by H. L. Smith of NRL also shows intense cracking in the resin parallel with the windings. Much noise accompanies such progressive failure. An attempt is being made at Aerojet to predict the failure pressure from the sounds and shocks emitted during hydrotest but thus far without success.

Unwinding failures have been typical of about one-third of all hydrotest burst tests in the Polaris program. In the future this is expected to be higher because of recent improvements in the end closures. The bad effects of mechanical damage can be somewhat controlled by proper choice of winding lay-up. When layers of circumferential and longitudinal windings are interleaved as in Figs. 13 and 14, the first stages of failure tend to be inhibited. One spectacular unwinding failure followed the tensile failure of a group of rovings in which knots had been tied. The appearance after failure is shown in Fig. 15. In this chamber, the circumferential windings were on top of the longitudinals.

Effects of time under load and of environment (usually moisture) are of practical importance. Trials have been made and considerable discussion has occurred. At this time it is clear that time under load does cause some loss in strength. However, the mechanism of this strength loss is uncertain and there are apparent conflicts in current test data which require critical review. Although the topic must be regarded as of considerable importance, we will not attempt detailed discussion in this paper. A paper by Petker (15) is a good source of information on the effects of moisture.

With regard to GRP pressure vessels in tension the major weakness factors are:

1. Unnecessary stiffness at the large adhesive joints.
2. An alignment of local stress elevation due to an abrupt change of stiffness, a poorly tensioned roving, etc.
3. Numerous fiber breaks due to abrasion, rough handling, etc. These may cause insufficient basic fiber strength even in terms



Fig. 13 "Unwinding" failure started intentionally from a mechanical cut in outer hoop windings of pressure bottles. These are 6-inch diameter internally pressurized bottles. The one on the left had a cut 0.0045 inch deep and the one on the right had a cut 0.010 inch deep. All of the hoop windings were outside. The effects of cuts have been reported elsewhere (Ref. 13).



Fig. 14 Unwinding failure developing under constant pressure with no cut



Fig. 15 Hydrotest failure started by knots in rovings in a large g.r.p. rocket chamber

of average strength of short lengths.

4. Lack of parallelism of individual fibers.

5. Poor toughness in shear fracture in the resin and between resin and glass.

6. Degradation due to time under load and environmental effects.

The major strength factors are:

1. Adequate protection of individual fibers from damage between drawing and chamber winding.

2. Careful use of flexible elements (elastomers, say) in reducing excessive stiffness and in reducing the stress concentration where a change of stiffness is unavoidable.

3. Parallelism of individual fibers and strands.

4. A substantial degree of flexibility and toughness in the epoxy resin.

5. A basic high potential fiber strength in the glass.

Information on glass fiber wound chambers loaded by external compression is entirely of the nonproduction, laboratory model type. When a cylinder constructed with multiple layers of glass cloth (or glass reinforced tape) is subjected to external pressure the failure mechanism initiates where the compressive stress is greatest. This means failure starts in the innermost layer. Two of the observed failure mechanism proceed as follows:

A. The first major nonlinear reaction appears to be parallel kinking of the glass fibers along a line and in a direction which allows greatest local relief from compressive stress in the layer. This is accompanied by debonding from the adjacent layer. With growth of the debonded area the adjacent layer is overstressed in compression and the innermost layer becomes flexible enough to buckle inward. The cycle of kinking, debonding, and inward buckling progresses in unstable fashion once it is well begun. Figure 16 shows the typical appearance of a resin poor filament wound cylinder after failure under external pressure. The failure was apparently progressive.

B. When the glass reinforcement is in the form of square weave glass cloth, a shear crack mode of failure has been sometimes observed. Figure 17 shows compression failure with a minimum of delamination in a cloth reinforced cylinder relatively rich in resin. It is believed a tendency toward kinking is again the leading failure element. Possibly due to the bending influence of the square weave pattern, fiber breaks and resin cracks are well enough aligned so that crack development with additional strain is relatively easy. Compression is relieved by a more localized degree of debonding than for (a) and by an inward offset of the layer on one side of the line of fiber breakage. After the affected region of the innermost layer ceases to carry compressive load the adjacent layer is overstressed. The imposed local shear strain due to the offset causes fiber breakages and offset in this second layer. Thus the shear crack develops layer by layer



Fig. 16 Typical inward buckling and debonding between layers of an externally pressurized filament wound cylinder. The failure is believed to have been progressive with much delamination



Fig. 17 Failure under external pressure of a cloth reinforced plastic cylinder. Note the local shear buckling of the cloth layers with a minimum of delamination

until a stress relieving offset is formed completely through the wall of the cylinder.

Although no detailed summary can be made at present of the strength and weakness factors of GRP chambers subjected to external pressure, it is clear that interlayer debonding weakened the structure failing by mechanism A. In the case of mechanism B, the initial kinking of the glass threads due to the square cloth construction was obviously detrimental. Somewhat more generally, one can recognize that these failures must start before they can occur and that methods of introducing a high degree of perfection in geometry and constitution of the innermost layers will substantially increase the compressive strength.

The feasibility of GRP structures for high external pressure service may well depend upon development of inner surface control schemes which accomplish the objective of delaying failure mode development and at the same time lend themselves to large scale fabrications.

REFERENCES

1. Kaplan, M.F., "Crack Propagation and the Fracture of Concrete," A.C.I. Journal Proc. V58, No. 5, pp. 591-610, Nov. 1961

2. Glucklick, Joseph, "The Flexural Static and Fatigue Behavior of Portland Cement Mortar," T. and A. M. Report No. 622, Aug. 1962

3. Brown, James H., "Elasticity and the Strength of Sea Ice," MIT Endicott Hours Conference on Engineering Glasiology, Dedham, Mass., Feb. 1962

4. Irwin, G.R., and Kies, J.A., "Fracturing and Fracture Dynamics," Welding Journal Research Supplement, Feb. (1952)

5. Goodyear Aircraft Corp. Final Report GER 10741, "Design Study to Improve the Structural Efficiency of Polaris Filament-Wound Motor Cases," Buweps Contract NOw 61-0500-C (FBM), 15 Aug. 1962

6. Ripling, E.J., Mostovoy, S., and Patrick, R.L., "Application of Fracture Mechanics to Adhesive Joints," Status Report to NRL on Contract Nonr-3544(00)(X) for Apr. 1962 to Aug. 1962

7. Irwin, G.R., "Relatively Unexplored Aspects of Fracture Mechanics," T. and A. M. Report 240, Univ. of Ill., Feb. 1963

8. Zender, G.W., and Deaton, J.W., "Strength of Filamentary Sheets with One or More Fibers Broken," NASA Tech. Note TN D-1609 (1963)

9. Hedgepeth, J.M., "Stress Concentrations in Filamentary Structures," NASA Tech. Note TN D-882 (1961)

10. Kies, J.A., "Maximum Strains in the Resin of Fiberglass Composites," NRL Report 5752, March 26, 1962

11. Schulz, J.C., "Maximum Stresses and Strains in the Resin of a Filament-Wound Structure," Aerojet-General Corp., Tech. Memo. 187SRP Dept. 4760, Apr. 1962

12. Schulz, J.C., "Maximum Stresses and Strains in the Resin of a Filament-Wound Structure," Society of Plastics Industries Proc. of the Technical and Management Conference, Reinforced Plastics Div., Sect. 7-D, Feb. 1962

13. Kies, J.A., Smith, H.L., Wolock, I., and Mast, P., "Vulnerability of Filament-Wound Pressure Bottles to Mechanical Damage," Proc. Conference on Structural Plastics Adhesives and Filament-Wound Composites, sponsored by Materials Lab., Materials Central in Dayton, Ohio, Dec. 11, 12, and 13 (1962)

14. Kies, J.A., "Prediction of Failure due to Mechanical Damage in the Outer Hoop Windings in Fiberglass Plastic Pressure Vessels," NRL Report 5736, Jan. 18, 1962

15. Petker, Ira, "Structural Materials Dev lopment Phase III Increment 7 Polaris A3," Aerojet Report of Progress for period 9-22 Feb. 1963

PROPAGATION OF HYDROMAGNETIC WAVES IN AN ATMOSPHERE OF ARBITRARY ELECTRIC CONDUCTIVITY

H. F. Ludloff
Pacific Missile Range
Point Mugu, California

(From work in cooperation with F. Reiche)

In the following, the interaction of time-dependent electromagnetic fields ("electromagnetic waves") with an atmosphere of arbitrary electric conductivity σ is investigated over the whole range of $\sigma(\sigma = 0 \cdots \infty)$; the range of small σ seems most important for terrestrial applications.

Owing to the existence of the Lorentz force $(1/c)(\underline{j} \times \underline{H})$, acting on each fluid element, the electromagnetic waves will generate fluid vibrations, which in this way can be transmitted over very large distances and produce far-reaching effects on the stability of flow patterns. Furthermore, it is also evident that, for small conductivities, when the gas behaves in some respects like a dielectric, the displacement current can no longer be neglected, as is usual. Therefore, the complete Maxwell equations have to be used, which are known to be Lorentz invariant.

II. EQUATIONS OF MOTION

Let us consider the cases in which an external magnetic field H_0 (e.g., the earth's magnetic field) and the direction of propagation of the waves are either parallel or perpendicular to each other. There shall be no external electric field E_0. Let us further assume that the perturbations of the flow field (due to the electromagnetic waves), and of the electromagnetic field (due to the fluid motion) are relatively small.

Hence,

Case I:	Case II:
$\underline{q} = (u, v, 0)$	$\underline{q} = (u, v, 0)$
$\underline{H} = (H_0 + h_x, h_y, 0)$	$\underline{H} = (h_x, H_0 + h_y, 0)$
$\underline{E} = (e_x, e_y, e_z)$	$\underline{E} = (e_x, e_y, e_z)$

Then, assuming that in our considerations $\epsilon \sim \mu \sim 1$ and that the charge density is everywhere zero, we have Maxwell's equations:

$$\text{curl } \underline{e} + (1/c)\dot{\underline{h}} = 0; \qquad \text{div } \underline{e} = 0$$

$$\text{curl } \underline{h} = (1/c)\dot{\underline{e}} + (4\pi/c)\underline{j}; \qquad \text{div } \underline{h} = 0$$

Ohm's law:

$$\underline{j} = \frac{\sigma}{[1 - q^2/c^2]^{\frac{1}{2}}} \left\{\underline{e} + \frac{1}{c}(\underline{q}\times\underline{H}) - \frac{q}{c}\left(\frac{\underline{q}}{c}\cdot\underline{e}\right)\right\}.$$

Conservation of mass:

$$\frac{\partial}{\partial t}\left(\frac{\rho_0}{[1 - q^2/c^2]^{\frac{1}{2}}}\right) + \text{div}\left(\frac{\rho_0 \underline{q}}{[1 - q^2/c^2]^{\frac{1}{2}}}\right) = 0.$$

Conservation of momentum:

$$\frac{\rho_0}{[1 - q^2/c^2]^{\frac{1}{2}}} \frac{d}{dt}\left[(1 + i_0/c^2)\frac{\underline{q}}{[1 - q^2/c^2]^{\frac{1}{2}}}\right] + \nabla p = \frac{1}{c}(\underline{j}\times\underline{H}).$$

Balance of energy:

$$\frac{\rho_0 c^2}{[1 - q^2/c^2]^{\frac{1}{2}}} \frac{d}{dt}\left\{\frac{1 + i_0/c^2}{[1 - q^2/c^2]^{\frac{1}{2}}}\right\} - \frac{\partial p}{\partial t} = \underline{j}\cdot\underline{e}.$$

Here ρ_0 is the "rest mass" density, i_0 the stagnation enthalpy, p the pressure, and c the light velocity.

We can linearize all equations of motion, neglecting second orders in the perturbation quantities; this greatly simplifies particularly the fluid equations. If we are interested in plane waves, all quantities can be made to depend only on x and t.

In case I, Maxwell's equations and Ohm's law yield

$$\frac{\partial e_z}{\partial x} = \frac{1}{c}\frac{\partial h_y}{\partial t} \tag{1}$$

$$\frac{1}{c}\frac{\partial e_z}{\partial t} + \frac{4\pi\sigma}{c}\left\{e_z - \frac{v}{c}H_0\right\} = \frac{\partial h_y}{\partial x}. \tag{2}$$

It can be shown that, after a short relaxation time

$e_x = e_y = 0$. Also, $h_x = 0$;

the fluid equations reduce to

$$\frac{\partial \rho_0}{\partial t} + \rho_0 \frac{\partial v}{\partial x} = 0 \tag{3}$$

$$\rho_0 (1 + v_0/c^2) \frac{\partial v}{\partial t} + a^2 \frac{\partial \rho_0}{\partial x} = 0 \tag{4}$$

$$\rho_0 (1 + v_0/c^2) \frac{\partial v}{\partial t} = \frac{\sigma \mu_0}{c} \left(e_z - \frac{v}{c} H_0 \right) \tag{5}$$

$$\rho_0 \frac{dv_0}{dt} - a^2 \frac{\partial \rho_0}{\partial t} = 0. \tag{6}$$

Here a is the sound speed. In Eq. (6), the term representing the electromagnetic dissipation has been dropped, because it is of second order.

III. SOLUTIONS IN CASE I

Now, assume waveform solutions

$$\begin{aligned} \rho_0 &= \rho_{00} + \rho_1 \exp(-i\omega t + ikx), \\ v_0 &= v_{00} + v_1 \exp(-i\omega t + ikx), \\ u, v, e_z, h_y &= u_1 \exp(-i\omega t + ikx), \quad \text{etc.} \end{aligned} \tag{7}$$

Since $k = k_r + ik_i$, we have, in general, spacially damped waves. Upon substitution of Eq. (7) into Eqs. (2)-(6), we get five homogeneous, linear equations for the five unknowns ρ_1, u_1, v_1, v_1, e_{z1}, whose coefficient determinant must vanish:

$$\begin{vmatrix} \rho_1 & u_1 & v_1 & v_1 & e_{z1} \\ \omega/k & -\rho_{00} & 0 & 0 & 0 \\ a^2 k & -\omega\tilde{\rho}_{00} & 0 & 0 & 0 \\ a^2 & 0 & -\rho_{00} & 0 & 0 \\ 0 & 0 & 0 & \dfrac{i\omega 4\pi\sigma H_0}{c^2} & -k^2 + \dfrac{\omega^2}{c^2} + i\dfrac{4\pi\sigma\omega}{c^2} \\ 0 & 0 & 0 & \sigma H_0^2/c^2 - i\omega\tilde{\rho}_{00} & \sigma H_0/c \end{vmatrix} = 0$$

Here $\tilde{\rho}_{00} = \rho_{00} (1 + v_{00}/c^2)$.

The determinant splits into two factors; the first factor yields

$$V_1^2 = \frac{\omega^2}{(k_r^2)_1} = \frac{a^2}{1 + v_{00}/c^2} = \tilde{a}^2; \quad (k_i)_1 = 0. \tag{8}$$

The second factor yields

$$\left.\begin{matrix}(k_r^2)_1\\(k_i^2)_2\end{matrix}\right\} = \frac{\omega^2}{2\sigma^2}\left\{\left[\frac{(1+p)^2+s^2}{1+s^2}\right]^{\frac{1}{2}} \pm \left(1+\frac{p}{1+s^2}\right)\right\}^{-1}, \qquad (9)$$

$$V_2^2 = \omega^2/(k_r^2)_2,$$

where

$$s = \omega\rho_{00}\sigma^2/\sigma H_0^2 \qquad \text{and} \qquad p = 4\pi\rho_{00}c^2/H_0^2$$

are dimensionless. If we vary s (that is, for example, σ), while p remains fixed, the phase velocity V and the spacial damping k_i are represented by the graphs in Fig. 1.

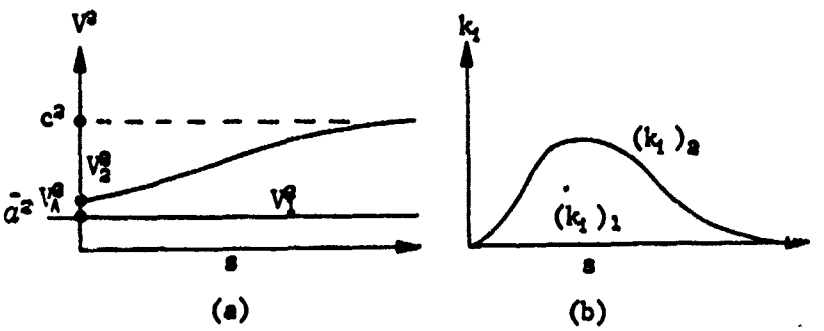


Fig. 1. Phase velocity V and damping k_i in case of parallel field.

In the limit $\sigma = \infty$, we obtain the generalized Alfvén speed[1] $V_A^2 = H_0^2/(4\pi\rho_{00} + H_0^2/c^2)$ and the sound speed $a^2/(1+\gamma_{00}/c^2)$; for $\sigma = 0$, we obtain the speed of light and speed of sound. The damping k_i for wave 2 is zero at $\sigma = 0$ and $\sigma = \infty$, although for different reasons. The sound wave 1 is always undamped.

The electromagnetic and the fluid vibrations have the same ω and k; but $\omega/k_r = V$, which varies with s. The polarization of wave 2 is transverse, that of wave 1 is longitudinal, that of the electromagnetic waves is still transverse.

The above expressions for the Alfvén speed and the sound speed differ somewhat from the customary ones; especially for very low densities, as in the outer atmosphere, the value for V_A would become unbounded, if the presence of the second term in the denominator would not keep it below c. This second term H_0^2/c^2 can be readily seen to be due exclusively to the displacement current, whose addition makes the Maxwell equations relativistically invariant.

[1] See P. Reichel, "Basic notions of relativistic hydromagnetics" (unpublished).

[Dropping $(1/c)(\partial e_z/\partial t)$ in (2) leads to omission of the term ω^2/c^2 in the last column of the determinant, and to the classical value of the Alfvén speed.]

From the last row of the determinant above, we may obtain an estimate of the amplitude of the fluid vibrations. In the limit $\bar{\rho}_{00} \sim 0$, we obtain

$$v_1/c \sim e_{z1}/H_0 . \tag{A}$$

This means, for example, that for the earth magnetic field and strong electromagnetic beams, the vibrational amplitudes can be very appreciable.

IV. SOLUTIONS FOR CASE II

Analogous considerations for case II yield the following system of linearized differential equations:

$$\frac{\partial e_z}{\partial x} = \frac{1}{c}\frac{\partial h_y}{\partial t} \tag{11}$$

$$\frac{\partial e_z}{\partial t} + 4\pi\sigma\left(e_z + \frac{u}{c}H_0\right) = c\frac{\partial h_y}{\partial x} \tag{12}$$

$$\frac{\partial \rho_0}{\partial t} + \rho_0\frac{\partial u}{\partial x} = 0 \tag{13}$$

$$\bar{\rho}_{00}\frac{\partial u}{\partial t} + a^2\frac{\partial \rho_0}{\partial x} = -\frac{\sigma H_0}{c}\left(e_z + \frac{u}{c}H_0\right) \tag{14}$$

$$\bar{\rho}_{00}\frac{\partial v}{\partial t} = \frac{1}{c}j_z h_z \sim 0 \tag{15}$$

$$\rho_{00}\frac{d\rho_0}{dt} - a^2\frac{\partial \rho_0}{\partial t} = \sigma e_z\left(e_z + \frac{u}{c}H_0\right) \sim 0. \tag{16}$$

After introducing wave form solutions, we obtain from (11) and (14) the following linear amplitude equations:

$$ike_{z1} + \frac{i\omega}{c}h_{y1} = 0 \tag{11a}$$

$$[4\pi\sigma - i\omega]e_{z1} - ickh_{y1} + \frac{4\pi\sigma H_0}{c}u_1 = 0 \tag{12a}$$

$$\rho_{00}ikv_1 - i\omega\rho_1 = 0 \tag{13a}$$

$$\frac{\sigma h_0}{c}e_{z1} + \left[\frac{\sigma H_0^2}{c^2} - \bar{\rho}_{00}i\omega\right]u_1 + a^2 ik\rho_1 = 0. \tag{14a}$$

From (11a) to (14a) follows the determinantal equation:

$$k^4 - k^2 \frac{\omega^2}{a^2 c^2}\left[\frac{\bar{\rho}_{00}}{\rho_{00}} c^2 + a^2 + i \frac{4\pi\sigma}{\omega}\left(\frac{H_0^2}{4\pi\rho_{00}} + a^2\right)\right]$$

$$+ \frac{\bar{\rho}_{00}\omega^4}{\rho_{00} c^2 a^2}\left[1 + i \frac{4\pi\sigma}{\omega}\left(1 + \frac{H_0^2}{4\pi\bar{\rho}_{00} c^2}\right)\right] = 0. \qquad (17)$$

The solution of this biquadratic equation leads to two values of the phase velocity, V_3 and V_4, and to the dampings k_{i3} and k_{i4}, which are graphed in Figure 2.

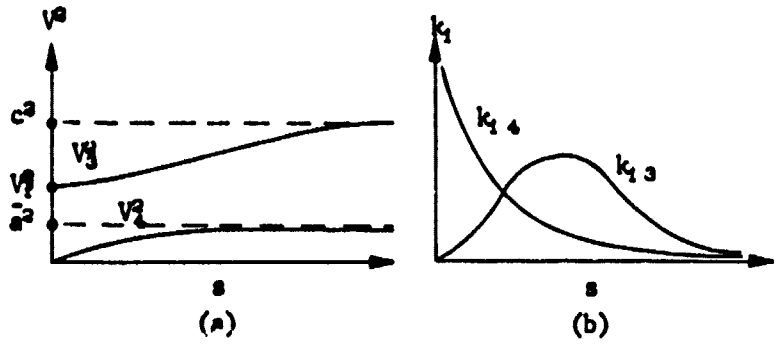


Fig. 2. Phase velocity V and damping k_i in case of perpendicular field.

From the determinant of Eqs. (11a)-(14a), it is seen that the polarizations of both fluid vibrations are longitudinal, while again the electromagnetic vibrations stay transverse. In the limit $\sigma = \infty$, the phase velocities are equal to the "fast wave" speed[1]

$$V_3^2 = (H_0^2 + 4\pi\rho_{00}a^2)/(H_0^2/c^2 + 4\pi\bar{\rho}_{00})$$

and to $V_4^2 = 0$. The corresponding dampings are $(k_i)_3 \sim 0$ and $(k_i)_4 \sim \infty$. It may be appropriate to elucidate the last statement.

From (17) it follows in the limit $\sigma \sim \infty$

$$(k_r)_4 = (k_i)_4 \sim \frac{1}{ac}\left\{2\pi\sigma\omega\left(a^2 + \frac{H_0^2}{4\pi\rho_{00}}\right)\right\}^{\frac{1}{2}}. \qquad (18)$$

From Eqs. (11a)-(14a), we obtain by elimination the following two relations between h_{y1}, e_{z1}, and ρ_1:

$$h_{y1} \sim -(k_4 c/\omega) e_{z1},$$

and upon substituting for k_4 from (18)

$$h_{y1} \sim \frac{\sigma^{\frac{1}{2}}}{ac}\{2\pi\omega(a^2 + H_0^2/4\pi\rho_{00})\}^{\frac{1}{2}} e_{z1},$$

so that

$$\lim_{\sigma=\infty} \frac{e_{z1}}{h_{y1}} = 0.$$

On the other hand, by another elimination process,

$$\frac{h_{y1}}{H_0} \sim \frac{\sigma^2 4\pi\rho_{00}}{H_0^2} \frac{\rho_1}{\rho_{00}}$$

so that

$$\lim_{\sigma=\infty} \frac{\rho_1/\rho_{00}}{h_{y1}/H_0} \text{ finite.}$$

In order to get the complete wave functions, we have to multiply these amplitudes by the wave factor $\exp\{-i\omega t + (i-1)k_r x\}$ where k_r is proportional to $\sigma^{\frac{1}{2}}$, according to (18). Hence, h_y and $\rho \sim \sigma^{\frac{1}{2}} \exp\{-\sigma^{\frac{1}{2}}x\}e_z$. Thus, at $x = 0$, h_y and ρ behave like step functions, which seems to indicate that this wave has the appearance of a contact discontinuity.

It is to be noted that for very large frequencies ω, the simple Ohm's law (see Sec. II) is to be replaced by a generalized one, which is based on a "two fluid theory"; in that case, as pointed out by Spitzer,[2] the following relation holds:

$$\frac{m_e c^2}{n_e e^2} \frac{\partial \underline{j}}{\partial t} = \underline{E} + \underline{q} \times \underline{H} + \frac{c}{e n_e} \nabla p_e - \frac{c}{e n_e} (\underline{j} \times \underline{H}) - \eta \cdot \underline{j}.$$

Clearly, this will lead to the appearance of the plasma frequency, and of the ion and electron cyclotron frequencies. It may be anticipated, then, that in the above-indicated curves of the phase velocity V and the damping k_i resonance values will occur. In other words, a dispersion theory will result.

V. GENERALIZATION

The preceding considerations have been generalized for arbitrary direction of plane wave propagation relative to a fixed magnetic field H_0.

If h_x, h_y, h_z and e_x, e_y, e_z are the induced magnetic and electric field strengths, and if u, v, w are the induced fluid velocities and ρ the resulting density disturbance, we may again assume solutions of wave form, e.g.,

[2] L. Spitzer, Jr., THE PHYSICS OF FULLY IONIZED GASES (Interscience publisher, Inc., New York, 1956), p. 21.

$$h = h_1 \cdot \exp\{-i\omega t + ik(n_x \cdot x + n_y \cdot y + n_z \cdot z)\}$$

and the amplitudes h_1, e_1, u_1, ρ_1 of these ten quantities obey the following ten linear equations:

$$ikn_y \cdot e_{z1} = i\frac{\omega}{c} \cdot h_{x1} \tag{23a}$$

$$-ikn_x \cdot e_{z1} = i\frac{\omega}{c} \cdot h_{y1} \tag{23b}$$

$$ik(n_x \cdot e_{y1} - n_y \cdot e_{x1}) = i\frac{\omega}{c} \cdot h_{z1} \tag{23c}$$

$$ikn_y \cdot h_{z1} = -\frac{i\omega}{c} \cdot e_{x1} + \frac{4\pi\sigma}{c} \cdot e_{x1} \tag{24a}$$

$$-ikn_x \cdot h_{z1} = -\frac{i\omega}{c} \cdot e_{y1} + \frac{4\pi\sigma}{c}\left(e_{y1} + \frac{v_1}{c} H_0\right) \tag{24b}$$

$$ik(n_x \cdot h_{y1} - n_y \cdot h_{x1}) = -\frac{i\omega}{c} \cdot e_{z1} + \frac{4\pi\sigma}{c}\left(e_{z1} - \frac{v_1}{c} H_0\right) \tag{24c}$$

$$-i\omega\rho_1 + ik\rho_0\{n_x u_1 + n_y v_1\} = 0 \tag{25}$$

$$-i\omega\rho_0 u_1 + ia^2 kn_x \rho_1 = 0 \tag{26a}$$

$$-i\omega\rho_0 v_1 + ia^2 kn_y \rho_1 = \sigma/c\left(H_0 \cdot e_{z1} - \frac{v_1}{c} H_0^2\right) \tag{26b}$$

$$-i\omega\rho_0 w_1 = \sigma/c\left(-H_0 \cdot e_{y1} - \frac{w_1}{c} H_0^2\right) \tag{26c}$$

These ten equations split into two systems: System I, relating e_z, h_x, h_y; u, v, ρ; System II, relating e_x, e_y, h_z; w.

System I yields for every direction $\bar{n}$ a "slow" and a "fast" wave velocity; System II yields the "transverse" wave.

System I leads to a complicated biquadratic equation in k^2, which for $\sigma = \infty$ and $\sigma = 0$ simplifies, to give the well known results of Friedricks-Kranzer and of propagation in a dielectric. For some, finite value of σ the "slow" and "fast" wave velocities can be shown to intersect.

System II leads to a single value for k^2; also here the values of the Alfvén speed for $\sigma = \infty$, and of $V + c$ for $\sigma = 0$ can be deduced.

SUMMARY

It is conceivable that electromagnetic waves, beamed into the upper atmosphere, where ionization and conductivity can not be neglected, produce fluid vibrations, which destabilize the flow pattern around incoming missiles. If the electromagnetic beams are sufficiently strong, the amplitudes of the air vibrations are appreciable, as shown by eq. (A).

ACKNOWLEDGEMENT

This investigation was supported, in part, by a grant of the National Science Foundation.