TRANSACTIONS

OF THE SEVENTH
SYMPOSIUM ON
BALLISTIC MISSILE
AND SPACE
TECHNOLOGY

VOLUME I

HELD AT THE UNITED STATES AIR FORCE ACADEMY, COLORADO, 13-16 AUGUST 1962
SPONSORED BY UNITED STATES AIR FORCE
AND AEROSPACE CORPORATION, LOS ANGELES, CALIFORNIA

AEROSPACE CORPORATION
POST OFFICE BOX 95085 • LOS ANGELES 45, CALIFORNIA
OSBORNE 5-4661

30 October 1962

To: Recipient

Subject: Transactions of the Seventh Symposium on Ballistic Missile and Space Technology, United States Air Force Academy, 13-16 August 1962

The Space Systems Division and Ballistic Systems Division, AFSC, and the Aerospace Corporation are happy to present to attendees of the Seventh Symposium and also to their organization libraries the first two volumes of the Transactions of that meeting.

Around the first of the year, the balance of the volumes of the Transactions--all of them classified--will be distributed, but only to organization libraries initially. Interested individuals should therefore check with their libraries for copies of the classified volumes.

Charles T Morrow

Charles T. Morrow
Chairman
Seventh Symposium

PREFACE

The Seventh Symposium on Ballistic Missile and Space Technology was held at the U. S. Air Force Academy, Colorado, on 13-16 August 1962. It was sponsored jointly by the Air Force and the Aerospace Corporation.

The six volumes of these Transactions contain both the unclassified and classified papers presented at the Symposium, and cover the categories as listed below:

I. Space Vehicles and System Design
Hypersonics and Re-entry Techniques
Propulsion
Structures and Materials
Mathematical Reliability Techniques

II. Earth Satellite Control and Control Systems
Control of Manned Space Environment
Simulation
Communications

III. Military Nuclear Rocket Systems
Advanced Propulsion Concepts
Special Techniques in Rocket Propulsion
Heat Transfer and Solid Propellants

IV. Guidance Development and Navigation
Earth Satellite Control and Control Systems
Defense Against Ballistic Missiles
Structures and Materials

V. Space Vehicles and System Design
Hypersonics and Re-entry Techniques
Communications
Radar Technology

VI. Midcourse Discrimination Techniques
Effectiveness of Multiple Re-entry Vehicle Systems
Use of High Altitude Nuclear Explosions as a Penetration Aid

Volumes I and II are unclassified; Volumes III through V are classified Secret, and Volume VI is classified Secret-Restricted Data.

PREFACE

The Seventh Symposium on Ballistic Missile and Space Technology was held at the U. S. Air Force Academy, Colorado, on 13-16 August 1962. It was sponsored jointly by the Air Force and the Aerospace Corporation.

The six volumes of these Transactions contain both the unclassified and classified papers presented at the Symposium, and cover the categories as listed below:

I. Space Vehicles and System Design
Hypersonics and Re-entry Techniques
Propulsion
Structures and Materials
Mathematical Reliability Techniques

II. Earth Satellite Control and Control Systems
Control of Manned Space Environment
Simulation
Communications

III. Military Nuclear Rocket Systems
Advanced Propulsion Concepts
Special Techniques in Rocket Propulsion
Heat Transfer and Solid Propellants

IV. Guidance Development and Navigation
Earth Satellite Control and Control Systems
Defense Against Ballistic Missiles
Structures and Materials

V. Space Vehicles and System Design
Hypersonics and Re-entry Techniques
Communications
Radar Technology

VI. Midcourse Discrimination Techniques
Effectiveness of Multiple Re-entry Vehicle Systems
Use of High Altitude Nuclear Explosions as a Penetration Aid

Volumes I and II are unclassified; Volumes III through V are classified Secret, and Volume VI is classified Secret-Restricted Data.

VOLUME I

CONTENTS

Addresses

Space Vehicles and System Design

Hypersonics and Re-entry

Propulsion

Structures and Materials

Mathematical Reliability Techniques

ADDRESSES

WELCOME ADDRESS

Maj. Gen. R. H. Warren
Superintendent
United States Air Force Academy
Colorado

The Air Force Academy is extremely honored that the Air Force and Aerospace Corporation selected the Academy as the site for the Seventh Symposium. All of you are invited to see as much of our grounds as you possibly can. We hope you will leave with a better understanding of our mission and our facilities.

Here are some basic facts about the Academy: The Cadet Wing reached its full strength of 2,500 this summer with the arrival of the class of 1966. The Air Force is requesting legislation to increase the Wing to 4,500. We have an annual input of 10,000 officers. Of these only 500 are Academy graduates.

Four classes have been graduated to date, and there are about 950 graduates on active duty. As a substitute for our lack of large numbers of alumni, we have two very active organizations that have supported us in many ways--the Air Force Academy Foundation and the Falcon Foundation. The former is about to present the Academy with a new 40,000-seat stadium.

Our $140 million facility here has but one objective--to turn out dedicated and motivated regular Air Force officers. The Academy will be completed, at least as originally planned, as soon as the Cadet Chapel is finished and accepted by the Government later this year.

The Air Force is very proud of the Academy's academic record during its brief history. Last year it ranked second among 187 important educational institutions whose graduates took a special graduate achievement examination. For the future, we hope to obtain authority to award master's degrees.

We need an air field for a flying indoctrination program. I consider this most important because the Air Force will place a major reliance on airplanes for a long time to come.

A field house for year-round football, baseball, and track is being considered, as is a guest house.

We hope to establish a visitors' center--a combination educational and historical facility to inform the public on the goals of both the Academy and

the Air Force. Last year there were over two million visitors to the Academy. This is the number one tourist attraction in Colorado.

In closing, I want to say that our facilities here are at your disposal; if there is anything we can do for you, let me or one of the Academy officers know.

WELCOME ADDRESS

Dr. Ivan A. Getting
President
Aerospace Corporation
Los Angeles 45, California

On behalf of Aerospace Corporation, I would like to welcome you to this Symposium and also to express to General Warren and his staff my sincere thanks for the assistance they are giving us. They have gone far out of their way and broken many of their traditional rules to help us.

My comments will be limited to two very short points. I can speak to you as a civilian scientist in a manner which perhaps would not be appropriate to some of my colleagues in uniform.

Let me point out as my first comment that, if you look back in history to the great inventions, the important scientific breakthroughs and technological strides, these often have been associated with nations in crisis. Count Rumford, born in New England as Benjamin Thompson, conceived a modern thermodynamic theory and determined the mechanical equivalent to heat while boring cannons in Munich when Europe was threatened by the Turks. More recently, the advances made in electronics and nuclear fission were also stimulated by a crisis facing our own Nation.

Today something like 70 to 80 percent of all scientific research and development in this country is sponsored by the Government. The bulk of it is for the Department of Defense. Scientists, therefore, cannot look upon these facts as a side issue. Our major role right now, whether we like it or not, is to work with the Government in running the scientific and technological race.

The second point I would like to make is that somehow there seems to be creeping into our public thinking the idea that anything peaceful is not connected with the military and that anything having to do with the military is bad. This dichotomy is dangerous. It is too simple. Historically, the military is a function of our Government; it is an arm to preserve our way of life--a means to preserve democracy.

The military has a peacetime role as well as a wartime role. The ships of our Navy visit harbors all over the world on disaster relief and other goodwill missions. They represent, as do our Air Force bases abroad, the United States Government and people. The Army Corps of Engineers, since the time of George Washington, has had a peacetime assignment in the construction of harbors and canals. The Navy for a long time has been a major source of astronomical data. It is also largely responsible for exploration of the ocean depths.

To preserve our democracy, it is important that we recognize the Department of Defense as both a military organization with a wartime mission and an agency of our Government and our people with a peacetime mission. The peacetime mission must be kept at a high level of performance to maintain fitness for wartime.

Recognizing that the importance of the speaker is often in inverse proportion to the length of his introduction, I simply present to you General Howell Estes, Deputy Commander of Air Force Systems Command.

WELCOME ADDRESS

Lt. Gen. Howell M. Estes, Jr.
Deputy Commander, AFSC, for Aerospace Systems
Los Angeles 45, California

The first half dozen of these seminars were jointly sponsored by Air Force research and development units for ballistic missiles and space on the West Coast and by our technical contractor--at the outset, Space Technology Laboratories, and now Aerospace Corporation.

In this time, ballistic missiles and space systems have become of a far more direct interest to the Air Force as a whole. For instance, since our last Symposium, large numbers of missiles have become operational at Strategic Air Command bases. Consequently, to recognize this broadened participation, this year your sponsors are the United States Air Force and Aerospace Corporation.

On behalf of the Air Force Systems Command as well as the entire Air Force, I would like to add my welcome to that already expressed by Dr. Getting.

As you know, our Symposium was previously held in Los Angeles. But this year we selected the Air Force Academy. Our choice was not based solely on the facilities available here. There were two other important considerations. First, this institution is the birthplace of the future Air Force. Second, that future Air Force will become more and more involved in science and technology. I think it is most appropriate, therefore, that our Ballistic Missile and Space Technology Symposium be held here. Furthermore, I believe it is important that you see the Academy and discover its objectives and capabilities.

Holding the conference here has certainly not been an easy task for either the local community or the Academy. The classes which have been away for the summer will start to return on Wednesday. I am intimately familiar with that fact, because one of the returning cadets is a son of my own, I am pleased to say. The community is a resort area and for it to provide for the fine attendance at the Symposium has been a real problem. I want to express our greatest appreciation to General Warren and his staff as well as to the Colorado Springs Chamber of Commerce and all the local organizations which cooperated with us so superbly.

Now to the Symposium itself. Over the years since 1954, we have built an extremely solid technology for ballistic missiles. The efforts of all concerned--the scientific and technical community and the United States Air Force--are not only commendable but truly amazing. The future of ballistic missiles seems largely to lie in additional applications and in updating our achievements.

Of importance is the fact that our space systems until now have rested entirely on ballistic missile technology. The task ahead is to develop a new technology especially appropriate to and totally associated with the space medium. In updating existing techniques and creating new ones, our foremost problem for the future will be that of rapid and effective exchange of scientific and technical information. It is my opinion that symposia such as this are a tremendous aid in meeting that problem. We have an excellent program for our Symposium this year, and I wish you all success in its implementation.

We are extremely fortunate today in our keynote speaker. Roswell Gilpatric was the first chairman of the board of trustees of Aerospace corporation. In his present post as Deputy Secretary of the Department of Defense, he continues his strong interest in the growth of technology. He has an extremely busy schedule, and we owe him our thanks for giving up some of his time to be with us.

This is the second time this year that he has keynoted a major symposium sponsored by the Air Force, the first being the Management Conference held last May at Monterey. That conference, from what we have heard from the participants, was extremely successful. No small part of that success was due to the keynote address delivered by our speaker.

At this time, then, let me introduce with the greatest of pleasure the Honorable Roswell Gilpatric, Deputy Secretary of Defense.

DISARMAMENT AND DEFENSE

The Honorable Roswell L. Gilpatric
Deputy Secretary of Defense
United States Department of Defense
Washington, D. C.

Someone asked me if this meeting, devoted primarily to ballistic missile technology, was an appropriate occasion to talk about disarmament and arms control. I said yes because there is no group more familiar with what modern weapons can do if they are ever set loose. No group has a clearer understanding of two requirements at stake. First is the need to work toward the day when it will no longer be necessary to create such weapons. Second is the need, so long as these weapons continue to be created, to minimize the chance that they will ever be used, and the extent to which they might be used should deterrence fail.

The great paradox we must live with is that, although we do not regard the prospect of nuclear war as anything but tragic, we cannot safeguard our vital interests unless we are prepared to accept the risk of just such a war. This is not a pleasant situation to be in. But the way of dealing with it is not to retreat to the simple-minded sloganeering of "better Red than dead," or the equally simple-minded "better dead than Red." Our aim, as the President has said, is to be "neither Red nor dead, but alive and free."

Our national policy is to maintain adequate military strength for the defense of our legitimate interests while seeking to lessen the international tensions which have made unavoidable the heavy level of armaments needed to provide that strength. At the same time we cannot sit by in the hope that the world will work out peaceful means of resolving its dissensions before nuclear war occurs. Hence, we must do everything possible now to reduce the risk of a war which neither East nor West could rationally want. There is nothing we can do to eliminate entirely that risk; there is much we can do to lessen it.

Some of the most important things we can do are not part of what is customarily thought of as disarmament or arms control. We do not, for example, think merely of numbers of weapons in estimating our defense needs. We think also of kinds of weapons. We spend more money assuring the survivability and control of our deterrent forces than in buying the missiles themselves. By putting these weapons underground, as with the hardened Minuteman sites, or beneath the sea in Polaris submarines, we assure that they cannot be wiped out by a surprise attack.

We believe this promotes stability in two ways. First, by assuring ourselves that we can absorb a surprise attack and still retain sufficient strength to strike back, we can with confidence wait until we are absolutely

sure we have been attacked before launching our own missiles. This condition reduces the chance of war by accident or misunderstanding. Second, by building missiles especially designed for a second strike, we lower somewhat Soviet concern that we may be tempted to strike first. Thus the very makeup of our strategic forces refutes, more strongly than words can do, the notion that we regard those forces as anything but a deterrent or defense against aggression. Force design is only one of the areas in which we have been working on unilateral steps to lessen the risk of war and its magnitude should it nevertheless occur.

Directly in the area of arms control and disarmament, the first task of our Defense Department, and of the defense establishments in other countries, is the technical one of evaluating the national defense implications of the various proposals that are being considered. For all of the governments involved have agreed that steps toward general disarmament must be designed so that no one side gains a significant military advantage from the disarmament process. Furthermore, all sides must have reasonable assurance that the agreed steps are being faithfully adhered to.

These requirements are not minor issues which can be brushed over in order to get on with the real business of disarmament. The real business is not disarmament as such but reducing the risk of war. It does not take much foresight to see that an improperly designed or inadequately controlled disarmament agreement could increase rather than diminish tension and danger. No nation should be put in a position where it, or some important segment of its leadership, believed that it could move aggressively to achieve a decisive advantage in world affairs. Nor would it be tolerable for another nation to feel its security so seriously threatened by an imbalance in the disarming process or by imperfections in the inspection procedure that it must repudiate the agreement and rearm.

This does not mean we should reject any disarmament agreement in which we perceive some risk, any more than we should accept any agreement in the name of disarmament without realistically assessing the risks involved. Rather we should seek to balance the risks involved in accepting or failing to accept each proposal. This approach leaves the extremists at both ends unhappy, with the result we may be called warmongers by one side and appeasers by the other. But between the alternatives of being called names or of blundering into a situation where either nuclear war or surrender is virtually inevitable, it is better to be called names.

The present balance of military power effectively eliminates the likelihood that any nuclear power would deliberately provoke a nuclear war. But we still live in danger. War by accident or misunderstanding, or folly of one kind or another, remains possible. Formal disarmament and arms control agreements are one, although only one, of the approaches through which we can work to reduce the danger inherent in the existence of modern weapons technology. So long as cold war tensions exist, that is, so long as the distrust and aggressive tendencies that first caused the arms race exist, the political leaders on both sides will necessarily look to their defense establishments for advice on the military significance of the proposals that are being considered. This alone clearly gives the defense establishments on both sides an important role in the area of disarmament and arms control.

It does not mean, however, that the defense spokesmen should hold a veto over the judgments of political officials. Political leaders must always make the final decision. Specifically, they must decide whether broader considerations of statecraft shift a balance of risk that may seem unacceptable from a narrowly technical, in this case military, point of view. But unless the political leaders have before them a carefully thoughtout evaluation of the technical considerations, they cannot reach a judgment on the broader balance of risks. That is because they do not know, with any confidence, how great or how small the technical risks are. Thus foot-dragging is almost bound to occur on one or both sides as negotiations progress.

The point here is not to argue that the defense establishments should be consulted, for this is being done here and in other major countries involved in disarmament negotiations. The point is rather that the responsible defense officials, military and civilian, and their scientific and engineering advisers, such as many of you here today, must come to know the special problems of disarmament and arms control. Only thus can they properly perform in this area and in the broader area of minimizing the chance of a nuclear war and of the damage that might be done should deterrence fail.

When we move toward general disarmament, the situation becomes particularly difficult. If the military power of the East and West can be thought of as two sets of chess men, there would be no problem in disarming at the beginning of the game. Each side would be exactly in parity with the other. By agreeing to remove a piece at a time, both sides could quickly disarm down say to the kings, representing the power necessary for internal police functions but insufficient to constitute a threat to the other side.

Once the game is under way, however, the problem becomes much more complex. Then it would not be so easy to agree on the order of removing the pieces so that at no point could either side have an advantage. The same pieces may have different values, depending on the strategic positions of the two sides. The two sides may no longer have exactly the same number and kinds of pieces. The existence of a mutual threat, as there is in real life in the form of a possible nuclear disaster, could provide some assurance to each side that the other had sufficient incentive to proceed with disarmament to counter the temptation to take advantage of any slight imbalance at a given stage.

But the real world is enormously more complicated than any problem set up on a chess board. There are not two sides, but two alliances with differences of viewpoint within each. There is also a third group of powers not allied with either side or with each other, and distrustful of both established blocs. There is a continuing technological revolution so that new pieces are appearing all the time. Still another complicating factor is the element of secrecy. In chess board disarming, each side at least would have full knowledge of whether each side is, stage by stage, fulfilling its commitments.

Disarmament in the real world would be far from easy, even without the problem of secrecy. With it the obstacles appear almost insurmountable. There are people passionately concerned with disarmament who take the view, "the Russians are obsessed with secrecy, the Americans are obsessed with the

danger of cheating, and between them they make progress almost impossible." A neat formulation but its implication that both sides are equally unreasonable is misleading. For the risk to Soviet national security in relaxing their passion for secrecy is trivial compared to the risk to the West if flaws in the verification process were to tempt the Soviets into failing to fulfill their commitments.

Russian suspicion of foreign intrusion is not a creation of the Soviet government. It is an attitude with deep roots going far back into the Czarist period. The Soviets, with some justification, regard this tradition of secrecy as a strategic asset. We have worked hard to devise proposals, to secure adequate verification with a minimum of intrusion into Russian affairs and with maximum safeguards against the possibility of misuse of the inspection system. We have sought, as the President stated a fortnight ago, to keep our position "squarely in line with the technical realities." We have attempted to meet the principle that the degree of inspection should be proportionate to the degree of disarmament. But, at heart, there is no getting away from the fact that disarmament and secrecy are inconsistent goals. The Soviets cannot, as Secretary of State Rusk has pointed out, "eat the cake of disarmament and keep the cake of secrecy." The Russians must decide whether they are really interested in disarmament, or whether they are really interested in secrecy.

I would also suggest that the passion for secrecy is far from an unqualified asset to Soviet security. To the extent that they agree with us that the national security of all countries is threatened by the danger of war, they should realize that secrecy has the unavoidable effect of increasing tensions and the arms race, and so adding to the risk of war. By simply reading material published openly in the American press, the Soviets know within quite reasonable limits what the United States is doing in the field of armaments. The West, on the other hand, must guess, to an important extent, what the Russians are doing. The situation is not wholly unlike that of a foot race with one runner blindfolded. Such a race tends to be all-out because the blindfolded runner can never be sure that he is not falling behind.

To understand the impact of Soviet secrecy, one must understand the dynamics of the arms race between the great powers. Much thinking about the present arms race is colored by historical experiences with arms races which were primarily open and quantitative. Known quantities of known armaments were built up. Today the arms race is still quantitative in a sense, but to a larger extent it is qualitative. As this audience well knows, it is a matter of continuous technological innovation, of moving into newer weapons systems more than multiplying older ones.

The Soviets can to a large extent react to what we are actually doing rather than to exaggerated fears of what we might be doing. But, unfortunately, so long as we have so little knowledge of what the Soviets are doing, we must base our preparations to a significant extent on what we think they are capable of doing. This is an important consideration in view of the relatively long lead-times required for the development and production of weapons systems. We cannot wait until Soviet weapons have been developed, produced and deployed in such quantities as to be evident before we begin our own cycle of development and production. Thus the Soviets are forced to

work hard to match the efforts that they know we are making to match the efforts that we think they are making.

It is too much to expect the Soviet Union will move swiftly to an open society. But perhaps they can be brought to appreciate the way in which their passion for secrecy not only makes disarmament extraordinarily difficult but also is itself an important cause of the arms race.

All of this does not add up to a cheerful outlook for disarmament, particularly for any immediate approach to the long-term goal of general disarmament. We can take comfort, nevertheless, in the increasing interest and understanding of the problems of disarmament and arms control which is apparent in this country and which is being paralleled in other countries. Likewise, we can reasonably hope for at least a beginning to progress in this field at the current Geneva conference.

Here I go back to the need for all of us in the defense establishments of the countries involved to play a useful and affirmative role in the field of disarmament and arms control. It is a role going beyond simply recommending a yes or no to various points on the proposals we are asked to consider. We must come forward ourselves with constructive proposals.

We must be asking ourselves not merely, "What can we stand in the way of disarmament and arms control without weakening our security," but "What can we suggest that will add to our security." This is an enormously difficult area. It is easier to think merely in terms of building ever stronger defense. But I have no doubt that if the defense establishments on both sides face up to the situation, arms control proposals can be developed which will add to the security of all nations without significantly jeopardizing the legitimate interests of any.

To be able to do this, the defense establishments on both sides have to make this problem area a major interest and concern. Reviewing the proposals the United States has made at Geneva has made us in the U. S. defense establishment increasingly conscious of the importance of military stability and of the part we should play in promoting that stability.

In some areas related to disarmament and arms control, I think we have been doing very well. We have worked out elaborate precautions to reduce the chance of war by miscalculation or accident. We have designed our systems in ways to make extremely unlikely the unauthorized firing of a nuclear weapon. We are trying hard to achieve the most reliable command and control systems, so that even in the event of a nuclear exchange there would be some hope of being able to limit the extent of the exchange and to minimize, to the extent possible, the damage to civilian targets. This is not to say that we have satisfactorily reduced the risks of the arms race. None of what we are doing is a substitute for disarmament, any more than parachutes are a substitute for adequate landing gear. But these are steps which, in important ways, help to reduce the danger. We hope that other defense establishments are moving along similar lines.

I would like to see all of those associated with defense establishments pay still more attention to these problems--the problems of stability and of

the arms race. The factors of stability and of the dynamic effect on the arms race should be considered in every decision we make, whether in the area of strategic doctrine, force structure, or research and development. This applies equally well to both sides. Indeed, the steps we can take unilaterally to minimize the dangers depend to some extent on what the other side may be doing. A decision to refrain from development of a particular weapon which might have a destabilizing effect, or which might stimulate the arms race, cannot be made without regard to what other nations may be doing. This is clearly the case, for example, with the possible spread of the arms race to outer space.

Finally, let me say a word about the hazards of slogans in the arms control area. A few weeks ago the President made a speech at Yale on the American habit of carrying on our domestic debate over economic policy in terms of myths that may never have been valid and which certainly cannot be validly applied to the American economy today. The same can be said about the problems of reducing the risk of nuclear disaster. This area is difficult enough without being confused by discussion carried on in terms of vaguely defined and sometimes meaningless slogans. For responsible governments--and the world is in a sorry fix if any of the governments with access to nuclear weapons are irresponsible--do not act on simple-minded slogans. To the extent that the arguments over arms control and disarmament are forced into terms of conventional myths and cliches, productive communication and exchange of views are inhibited.

To pursue at the same time policies of both armament and disarmament is not contradictory. No nation is going to go far on disarmament unless it considers disarmament policy frankly in conjunction with armament policy and the reasons it feels it should pursue a given armament policy. Both are part of the effort to provide for the national security. To be sure, as Mr. Justice Frankfurter has pointed out, "To do the contradictory is a tough problem; namely, with might and main to try to prevent from coming to pass the very situation in relation to which you're making preparations. That's what we have to do in life so much. You build a fireproof house and nevertheless take out fire insurance."

THE SPACE CHALLENGE

The Honorable Joseph V. Charyk
Under Secretary of the Air Force
United States Air Force
Washington, D. C.

As I scanned the program for your meeting and noted the wide range of technical subjects covered, I could not help but be impressed with the scope, breadth, and dynamic character of our national program in missile and space technology. In turn, I sought to reflect on this total effort--the nature of the challenge, the form of the response, the strengths and the weaknesses, and, on the most mystic item of all, the future. It is a difficult thing for one in Washington to do; the close proximity to the day-to-day successes, the failures, the hopes and the frustrations makes it perhaps not the best observation post for an objective view. But I hope you will forgive me if I indulge for a few minutes in this dangerouse pastime. I believe it is essential for us all to take stock from time to time and to review appropriate areas of concern; for in the melding pot of scientific, industrial, and governmental ideas lie the strengths of our democratic way of life.

It seems strange that so often we appear to overbalance; a particular management approach, idea, or technical field of endeavor becomes attractive and we swarm to embrace it--only to overbalance. Then the pendulum hits the opposite stop or the charge goes off on another tangent.

An impulse is introduced by the forces of the cold war and the reaction may be similar. In some respects, a system with such dynamic behavior can be effective if the inertia to change is not too great, if the system has reasonable damping, and if the necessary components are present. It becomes important to insure that these conditions are met.

Let us look then at the missile and space picture. Only eight years ago a serious effort in ballistic missile technology was launched. Today, the results are truly profound. Missiles of a wide variety of sizes, capable of global ranges and incorporating cryogenic, noncryogenic, and solid propellants are either in our operational inventory or soon will enter this inventory. The accuracies and payload capabilities are well beyond even the optimist's predictions of that time. From the management side, a new concept was introduced. The combination of technical skills required for this undertaking was such that the use of a system contractor in the classical sense was deemed unsuitable and the systems engineering contractor and associate contractors technical management scheme was born.

Because of the vital importance of the program, time was critical. Streamlined approval channels were established; broad authority was given directly to the program director; contractual actions were initiated in the

interests of expediting the activation of elements of the program, without serious consideration to performance incentives or penalties for nonperformance and to the contractual provisions therefor; multiple approaches were initiated on most critical items; maximum concurrency was the keyword. The missiles in the operational inventory today, and those soon to enter the inventory and capable of performance well beyond the initial goals, are adequate testimonials that the race to beat the clock was won.

In this case, we failed to move out on the ballistic missile program with sufficient vigor at an early enough date. Faced with the Soviet missile progress, maximum speed in our program was of the essence. The cost, of course, in dollars, was much higher than it would have been for a lower concurrency, longer paced program. In this case, there is no doubt in my mind that there was no other choice. But therein was planted the virus of what I might call the maximum concurrency disease, and it does not necessarily follow that the particular management approach used and the concept of maximum concurrency are the unquestionable answers to system development problems. Quite the contrary. Unfortunately, we have in our military space program at least one or two instances where a similar philosophy has left us with high expenditures and no operational capability. A complete and expensive concurrency total system development plan for a speculative, and perhaps mystic, military operational requirement is not the road to national strength in this important field--it could well be the road to dissipated resources, abortive programs, loss of national prestige, and most serious of all, if uncontrolled, a threat to national survival.

And now so that I may be equally unpopular with another school of thought, let me say a few words against the "sequentialists." This is the "orderly minded" faction that favors a complete series approach and a full demonstration of all of the elements of all the subsystems before proceeding on a step-by-step basis through all the other phases of a program. In my opinion, this is a sure road to the guaranteed obsolescence of the weapon systems which may finally emerge. With the pace of modern technology, and with the criticality of superior military capability to national security, to be second may well be suicidal. It is a risk we must not, and dare not, take.

The key then is a combination of simple, but difficult, things--a mature judgment as to the areas of critical importance, a willingness to take calculated development risks, a solid determination and will to terminate promptly or reorient programs where new factors have changed the tenets on which the program was launched, an assertive and prompt response to new and challenging possibilities, a willingness by industry and government to experiment with new contracting and management methods with premiums for performance and penalties for malperformance. We must abandon the "can't lose" contracts, the overstaffed technical and administrative layers of divided responsibility, the tonnage of formalistic paperwork, and return to the prompt reactivity, the willingness to accept clear cut responsibilities, and the eagerness to risk the penalty for failure for the chance of greater gain through performance. These are American traits and in our defense business they are today of crucial importance.

I have spoken on this theme before and I am delighted that now we are beginning to get some early indicators of what can, in fact, be done. An

interesting case is the Agena-D, a program to seek a standard upper stage vehicle for many of our space programs in the interests of achieving a respectable and vital reliability figure, a feature which was not a noteworthy characteristic of its early versions or, for that matter, not a noteworthy characteristic of a large number of our satellite vehicles.

The initial estimate in the early fall of 1961 was that the best possible schedule would result in an initial vehicle delivery in November of 1962, with initial launch in January of 1963. This was on the basis of a high-level effort, using procedures and techniques similar to those in effect for most of our programs. This was deemed to be completely unsatisfactory and a Special Group was activated to investigate ways and means by which the program could be expedited. Amazingly enough, after a thorough scrubbing of the program, a new schedule was evolved calling for initial vehicle delivery in April 1962 and initial launch in June 1962, a telescoping by about a factor of two.

The management and organizational arrangements called for far fewer people, simple reporting procedures, and an incentive-type contract based on schedules, performance, and cost. Both of the dates were met and the flights to date of the new vehicles have been successful.

In this case, I am confident that streamlined procedures accomplished--with streamlined management and attendant responsibilities, and with far fewer people, far fewer dollars, and far less time--a far better result than could have been accomplished under more normal rules of program development. In this case, however, the job was well specified initially and changes were not permitted. In an advanced development activity, however, and even assuming that the mistake of premature system development is not made, complete definition "by definition" is impossible and different ground rules must apply.

Another danger which stands out in my impression is the deadly peril of overlapping, divided and ill-defined management relationships. This is the sure road to the black plague of development programs. It can be so beset with further monitoring and checks and balance subsystems, purportedly to insure meshing, that a Gordian knot is assured.

I believe we have made much progress in attacking some of the basic problems to which I have alluded but, between industry and government, we still have a long way to go and it will take our best combined efforts to produce the kind of result that this great country of ours demands and deserves.

The improving trend that has recently characterized our programs, and our ever growing experience with the peculiar problems of space technology, give high confidence that with proper vigor and determination the necessary job can be done. Our success record in achieving orbit has gone from a low point of twelve failures out of seventeen tries in 1958, through eleven failures in thirty-nine tries in 1961, to a record of only one failure in sixteen attempts. This is dramatic progress and I think the experience that we are gaining, the techniques we are instituting, and the management responsibilities which are being assumed tend to insure that this quantitative improvement will continue.

There is one commodity that is common in all that I have said, one that, in a sense, is plentiful and yet one that cannot be bought for any price. That commodity is the human being, and in this context the competent visionary manager, the imaginative engineer, the skilled proud technician, the dedicated military man, and the inspired scientist form the links in the chain from which our national prestige and security is welded. And so it is appropriate and fitting, I believe, to dwell further on this key theme but, in particular, on this problem as it affects the Defense Department and its activities.

I am greatly concerned that much of our best technical and management talent is engaged in program acquisition rather than program execution. Our contracting methods, and their emphasis on the impressive paper proposal, have encouraged this unfortunate trend. Its reversal is not accomplished overnight but I believe progress is being made. With the development of suitable yardsticks, and the resultant measurement of performance in other programs, a more quantitative rating can be given to a particular contractor's record and this, in turn, used as a major determinant in selection on a new program. The contractor's record of retention of key personnel on going programs, or his early rejuggling to try for new business at the expense of the acquired programs, provides an important barometer as to his sense of obligation and dedication to the best fulfillment of his commitments. In cases where a contractor's performance degrades seriously, we have applied in the past, and will apply more broadly and stringently in the future, bars to invitations to him to compete for new business. The country needs results, not promises; it needs fulfillment of performance milestones, not the filling and filing of massive reams of proposal material; it needs the competent manager, not the soothing syrup dispenser in Washington. More and more we need the simple answer and not the ingenious all-function mousetrap; we need the clever solution, not the brute force answer; we need the original idea, not the bigger and better version of the old; we need realism, not fantasy; we need the identification of the critical elements, not the barrage approach. Our ability to put these simple maxims into practice will be a major factor in determining our place in the military technology role; there is no prize for second place but there is a penalty of the highest type.

A great responsibility rests also, of course, on those in government, and upon their decisions and judgments rest, in primary fashion, the future status of our military strength. It is disturbing and alarming, therefore, that the President's Special Assistant for Science and Technology should have to say recently, with candor, that the government's scientific ability has been deteriorating; that incompetent government scientists are the main reason so many of our programs go badly and that the situation must be improved quickly. The problems we face in the Air Force are prodigious. The AFSC spends about ten cents out of every tax dollar. The programs it manages represent the biggest engineering endeavors ever undertaken; the missile program will cost about five times more than the Manhattan project--it includes the biggest construction program ever undertaken. The country owes a great debt to those dedicated men in uniform who are managing this momentous undertaking. I have never met a more conscientious, dedicated and hardworking group of individuals. But, as you well know, powerful economic and other factors are daily at work and key experienced officers are lost to the service at a serious rate. Perhaps even more alarming for the future are the statistics for young

scientific and engineering officers of five years service or less. Over 90% leave at the end of the mandatory three-year period and 25% of the balance are lost by the fifth year.

Our new weapon systems demand ever increasing numbers of skilled officers and technicians. For example, a squadron of F-80's required seven electronic technicians. An F-106 squadron requires 70 electronic electricians, with much better and longer training and the attendant additional costs. A B-52 wing requires 136 technically trained officers--an Atlas missile wing six times as many. Communications and electronics personnel requirements are estimated to go up by about 15% in the next three to four years. We have too many older officers approaching retirement age and too few young men with long careers ahead of them. We have about 111,000 line officers now. Of these, 32,000 have 18 or more years of service; only 24,000 have less than five years service. This is one third less than is required to maintain the force at the present retention rate.

With respect to educational background, only 49% of our officers hold college degrees. If we exclude medical officers, chaplains, etc., this drops to 44%, a wholly unsatisfactory percentage and completely out of balance with the sophisticated requirements of the space age. This average must be raised and the retention rate must be improved. On an economic basis, it costs us, for example, at least $48,000 to put a man through the Air Force Academy. Advanced training at the Air Force Institute of Technology may cost another $30,000. The man will now have a master's degree in a field of high demand, will probably be in his mid-twenties, and will be a prime target for industry. Any business facing our costs and turnover would be in serious trouble with the stockholders. Yet too many of our stockholders--the citizens of the United States--appear to be unaware of the problem.

Skilled enlisted airmen are essential to our weapon systems. In critical skills like electronics, the retention problem is extremely severe. The man's first enlistment period is four years--after training and initial experience he is probably fully effective for only one year out of the four. The training costs for this one year of service are of the order of $25,000.

Retention, thus, is essential if we are to have an efficient military force, and, in the push-button age, the premium is higher than ever and, more than ever before, the trained man is the most precious commodity of all.

And so, having sought, very briefly, to touch on some of the critical problems we face and on some of the lessons we can draw from our experiences thus far in the missile and space age, let me speculate for a moment on the future.

I suspect that as seems so often the case, we will be overoptimistic on what can be done in the near future and overconservative on the capabilities which may be possible in the more distant future. We tend, in approaching a new field such as space, to associate it with the millenium for the extension of ideas and concepts for military capabilities to which we have grown attached--the orbital bomber, the maneuverable interceptor, the high ground observation point. But I have a feeling that the most exciting and vital

utilization of space will come from new technology that is peculiarly adapted to space and that will yield capabilities for which there are no parallels in our present systems. Let us not try to foist our classical concepts onto space, but let us exploit the new medium for what it offers uniquely. Let us not propose elaborate systems for doing military tasks in more expensive, elaborate ways. Let us seek rather to gain an understanding of the new medium, advance those technologies which exhibit unique characteristics in space, explore concepts based on such advances, and determine the kinds of military jobs which space makes possible. And perhaps the most vital point of all--we must not be constrained in such exploration by a requirement to justify an operational system requirement before approval to expend funds. This road leads either to foolish requirements, dissipated resources and wasteful programs aimed at ridiculous systems, or to no program at all. This road has been traveled before. The most costly saving is the money that was not spent on the good idea in good time. The most wasteful expenditure is the money spent on the shotgun approach when the panic button is hit. And, in this connection, I would like to emphasize as strongly as I can that the contest in which we are engaged has vital psychological and political elements--it is in truth a cold war. Space has inherent in it tremendous cold war implications. We must be ever alert and capable of responding to the type of challenges that may be hurled at us and which can do untold damage to prestige and morale if we are incapable of response. Thus, we must, in our military space programs, develop capabilities which can permit us to react promptly and effectively even though the particular challenge might not be considered a matter of decisive military importance.

As the technical leaders in the missile and space technology field, your performance, your sense of responsibility, and your judgment are critical elements in the determination of our military posture tomorrow. Exercise those responsibilities with utmost gravity, for on our military posture rests the future of the world.

SPACE VEHICLES AND SYSTEM DESIGN

PRELIMINARY LOCOMOTION ANALYSIS OF LUNAR SURFACE VEHICLES

W. B. Sponsler
Northrop Space Laboratories
1111 East Broadway
Hawthorne, California

ABSTRACT

This paper summarizes a preliminary study of vehicular mobility on the lunar surface. Basic modes of controlled locomotion are discussed, and the most promising (rigid and flexible wheels and tracks) are analyzed. The effects of surface characteristics, vehicle weight and tread configuration are related to traction, rolling resistance, range, obstacle capability and acceleration. Typical examples are compared, varying the lunar surface parameters to develop a relative evaluation and to present a method for further analysis. The necessity for slow speeds and the effects of lunar environment upon traction, ride, stability, braking and other factors are also discussed.

The work described in this paper was performed as a company-sponsored project.

PRELIMINARY LOCOMOTION ANALYSIS OF LUNAR SURFACE VEHICLES (U)

W. B. Sponsler
Northrop Space Laboratories
1111 East Broadway
Hawthorne, California

INTRODUCTION

We are standing on the threshold of a new technology, the development of extraterrestrial surface vehicles. Present concepts and experience in the design and use of locomotion devices for earthbound surface vehicles may be applied to lunar surface vehicles only with considerable caution. The weight limitations, system integration, operational environment and reliability requirements for lunar operation place the lunar surface vehicle in a realm all its own. One of the more important phases of this task is the selection of the locomotive system.

The problem of selecting the optimum locomotion modes for Earth vehicles has been a challenging one. Since the invention of the wheel, off-road locomotion development has been by "cut and try" methods, due to the complexity of the phenomena involved. Only within the last few years has a theoretical method emerged. The unknowns of the lunar environment such as surface factors, temperature effects, radiation and meteorite hazards, plus the uncertainties of communication, power supply, life support and mission requirements as they affect locomotion, only serve to increase the magnitude of the problem.

This paper summarizes one phase of our continuing study of mobility on the lunar surface. The work was based on procedures and techniques currently accepted by authorities in land locomotive sciences, and emphasizes a general comparative approach rather than a specific design. It forms a prerequisite for more detailed studies and finite designs later.

Design Boundaries

For purpose of this analysis, the following considerations were examined:

Mission Objectives

- Period: 1968 to 1980
- Location: near side of Moon on relatively easy terrain for first generation vehicles

- Range: 50 to 1000 miles without refueling
- Velocity: 1 to 10 mph
- Desired operation: lunar day and night
- Possible uses: exploration, construction, logistic, rescue and emergency shelter.

Lunar Environment

- Hard vacuum: $\sim 10^{-14}$ mm Hg
- Surface temperature: $\pm$ 250°F
- Gravitational acceleration: one-sixth of value for Earth
- Average slopes: less than 15 deg
- Quantity and size of craters, crevices and protuberances: unknown
- Surface soil: assumed to be of a dry frictional nature; strength and density unknown.

Mode Comparison

The basic modes of locomotion were investigated to permit selection of the most promising locomotion devices. Following is a summary of findings:

1. Crawling (similar to inch worm): Inefficient and mechanically unreliable due to complex mechanism required.

2. Sliding (snake-like): Inefficient over dry surfaces due to surface friction. Reliability decreases as the number of rigid body sections increase, while mobility increases as number of sections increase.

3. Jumping (grasshopper or "Oberth's Moon Car"): Takes advantage of lunar gravity. However, unreliability of mechanism, instability during free flight, unknown landing conditions, shock loads, and difficulty of testing prevent further consideration at this time.

4. Walking (man or horse type motion): Most advantageous for extremely rough terrain for short distances and light loads. Mechanical reliability and control difficult due to the complex mechanism. Efficiency on relatively smooth, level surfaces poor compared to the wheel. May be necessary for certain lunar areas.

5. Rigid wheel on rigid surface (train type): Most efficient man-made locomotion. If the lunar surface is hard and smooth, this is the ideal system.

6. Rigid wheel on elastic surface (wagon type): Rolling resistance is a function of vehicle weight, diameter and width of wheel, as well

as surface factors. Mechanical reliability, weight, and efficiency make it attractive if surface permits. Further study required.

7. Elastic wheel on rigid surface (auto on highway): Rolling resistance and turning resistance greater than rigid wheel. Ride is much improved on an unsprung vehicle. Reliability less than rigid wheel.

8. Elastic wheel on elastic (soft) surface (jeep off road): Rolling resistance low, steering easy, mechanism simple, shock attenuation good (unless tire too stiff). Traction a function of soil characteristics: in dry, frictional soil not much better than a hard wheel, neglecting slip. Reliability low if fabric used due to harsh environment, particularly at night. Further study required.

9. Track laying vehicle on rigid surface: Mechanical efficiency poor. Steering energy wasted in slewing track. Mechanical reliability poor. Not desired for hard level surfaces unless necessary to cross many depressions or crevices.

10. Track laying vehicle on elastic (soft) and rough surfaces: High flotation gives low rolling resistance, minimum slippage and fewer ground clearance problems; however, track is heavier, complex and has less reliability than wheel. Turning resistance high, rough ride over raised obstructions, greater dust and temperature problems. Further study required.

11. Rocket-powered flight (100 per cent powered or impulse): Control difficulties, dust problem and very low efficiency relegate this method to extremely rough terrain or very long distances.

It is expected that first landings and surface operations will be performed on the easiest lunar areas where wheel or tracks with their inherent advantages such as reliability, efficiency and ease of control may be employed. Therefore, this analysis for first generation vehicles has been directed toward wheel and track mobility.

TRACTIVE THRUST ANALYSIS

The first point of consideration is the vehicle tractive thrust ability. A self-powered vehicle must develop a thrust force at the ground equal to or greater than the external forces resisting movement for motion or acceleration to take place. If the surface cannot develop sufficient horizontal reactive thrust, the wheels or track will slip or slide excessively. Soil thrust may be classified in two categories: first, where surface is hard and does not deform under the load and, second, where it is weak (soft) and deforms under a load.

Concrete could be considered the ultimate in hard surfaces with hard-packed soil loaded below its yield as the other extreme of hard soil. The maximum thrust H generated on a hard surface is equal to the coefficient of

friction μ between the wheel and surface times the normal (vertical component of weight) force W, or

$$H = \mu W \quad . \tag{1}$$

The coefficient of friction for soil and dry concrete generally varies between 0.5 and 1.0 (2), depending on the materials and the surface roughness. If the lunar surface were dry compacted dust and the wheel tread had a frictional surface, a coefficient of friction of 0.7 or more could be expected.

Lunar dust may be hard and densely compacted as a result of the vacuum environment, as well as from vibrations from moon quakes and meteor impacts. Further, sputtering of the surface due to solar radiation may have produced a strong, sintered surface composed of meteorite dust and metallic (heavy) atoms (3); sputtering may also have pushed material down the slopes and filled in holes and crevices. Thermal shock, radiation effects, electrostatic action or some unknown phenomena may likewise have caused the surface to be weak and uncompacted.

If a hard, smooth surface exists and the vehicle does not break through, or if only a shallow dust layer exists, then rolling resistance will be low, slippage would be negligible and traction high. A rigid or semi-rigid wheel may be favored on this surface for efficiency and reliability.

On weak or soft surfaces, the maximum thrust is limited by the internal strength of the soil, rather than by the coefficient of friction between the tread and the ground. A wheel or track loaded with a vertical force W, produces a soil thrust H dependent upon the horizontal shearing strength of the soil. This soil strength consists of a cohesive strength as in wet clay and a frictional strength as in dry sand. The maximum ideal thrust which any particular tread can develop is given by (4)

$$H_{max} = A\,c + W \tan \phi \tag{2}$$

where

A = horizontal areas in contact with ground

c = coefficient of soil cohesion

ϕ = angle of friction (between soil grains)

W = normal force (weight of vehicle).

On Earth, c varies from 0 in dry sand to 3 psi in plastic-saturated clay, while ϕ varies from 0 in plastic-saturated clay to a maximum of 35 degrees in dry sand. If we assume that the lunar soil is a dry frictional type similar to loose dry sand, then

$$H_{max} = W \tan \phi = 0.7\ W \quad . \tag{3}$$

Thus we see that the tread area A is unimportant in a dry, frictional soil. Only the vehicle's normal weight component determines the maximum ideal

thrust, and since vertical force on the moon $W_{\leftmoon}$ is but 1/6 of the force on Earth $W_{\oplus}$, the maximum lunar thrust will be approximately 16 per cent of that on Earth.

Cleats or grousers have been found to increase thrust in cohesive soils but are of little value in frictional (dry sand) soils except for possibly digging through soft shallow soil to reach a stronger subsurface. Space link type treads (5) should be tested to assess their part in lunar operations. They may provide sufficient thrust advantages to be worthy of trade-offs in complexity, weight, rolling resistance and tread width.

The maximum thrust that a vehicle can develop on soft soil is actually dependent upon the optimum slip between the tread and the surface. Slippage is dependent upon the soil characteristics and the tread configuration (6). As the soil is horizontally compacted during tread slippage, its shear strength is increased. Some brittle soils, such as silt, produce maximum thrust at very low slippage after which their strength falls off rapidly. Here a short track or wheel may suffice. Other soils, such as loose sand, require more slippage to produce their maximum thrust. In this case, a longer track is more efficient for straight line travel, as it develops the required compaction with a lower percent of slippage.

During turning maneuvers soft pneumatic or wide wheels and tracked vehicles are very inefficient due to tread scuffing. The greater the ratio of the track length to the vehicle width, the greater the turning losses.

RESISTANCE ANALYSIS

The forces opposing the vehicle's thrust consist of rolling resistance (internal or external), obstacle resistance, inertial resistance (rotational and linear acceleration) and external work performed (towing, dozing, etc.)

On hard surfaces the major portion of the resistive force is actually internal, such as bearing friction and wheel deflections, with very little loss due to external surface deformation.

The rolling resistance R_h of a rigid wheel on a hard, smooth surface is given by (2) as

$$R_h = f \frac{W}{r} \tag{4}$$

where f (.002 to .020) is the coefficient of rolling resistance and r is the wheel radius in feet.

For example, let

$f = .011$

$r = 3.37$ ft

$W = 10{,}000$ lb for Earth

$W = 1667$ lb for the moon.

Then

$$R_h = 5.5 \text{ lb}_{\leftmoon} \quad \text{or 0.3 per cent of } W_{\leftmoon}.$$

The unit coefficient of rolling resistance f_t of a pneumatic wheel on a smooth, hard surface is due to carcass deflection resistance and is found by experience to average from .010 to .020 lb/lb of weight on auto and truck tires (2) with the higher values common on soft, low-pressure applications where greater carcass flexing occurs. For this use,

$$R_h = f_t W. \tag{5}$$

For example, let

$$f = .020 \text{ lb/lb}$$
$$W = 10{,}000 \text{ lb}_{\oplus} (1667 \text{ lb}_{\leftmoon}).$$

Then

$R_h = 33.3 \text{ lb}_{\leftmoon}$ which is 6 times the rigid wheel's resistance.

The hard surface rolling resistance of a tracked vehicle is also dependent upon its internal losses in bearings, links, track tension, etc, so that

$$R_h = f_m W. \tag{6}$$

Experimental results for values of f_m from commercial earth movers and Army vehicle tests may not be valid since these vehicles were designed for different surface and operational conditions and are usually very heavy. The "Hawk Missile Loader" (7), a small, lightweight, tracked vehicle, may more nearly approach the concept of a tracked lunar vehicle. Rolling resistance tests of this vehicle with clean parts, straight slow travel on smooth level surfaces, gave a unit resistance (f_m) of .040 lb/lb of vehicle weight on concrete and .045 lb/lb on hard, dry dirt.

With

$$f_m = .04 \text{ lb/lb and}$$
$$W = 10{,}000 \text{ lb}_{\oplus} \quad (1667 \text{ lb}_{\leftmoon}),$$
$$R_h = .040 \times 1667 = 66.7 \text{ lb}_{\leftmoon}$$

is twice the pneumatic or 12 times the rigid wheel's resistance.

It should be remembered that these internal resistances are based on tests on Earth and on the assumption that mechanical losses on lunar vehicles will be reduced in proportion to the vehicle's moon-to-earth weight ratio. This resistance ratio may be difficult to maintain at such low loads, especially with the stringent lubricant and seal requirements imposed by the harsh lunar environment.

A vehicle's rolling resistance on soft surfaces consists of two major components: the force expended on deforming the soil, and the internal losses in the wheel or track. This internal loss usually equals the previous hard surface resistances. There are several other losses: the bulldozing of soil in front of the tread, and the drag losses along the sides of the tread. These are usually low values and are neglected at this time.

The external resistance increases as the vehicle sinks into the soil as follows: when the vertical load exceeds the yield strength of the soil, the deflections and soil displacements become permanent, resulting in work being lost to this soil deformation. The force required to accomplish this work is equivalent to the rolling resistance R_s, and for a rigid wheeled vehicle is equal to: (5)

$$R_s = \frac{1}{(3-n)^{\frac{2n+2}{2n+1}} (n+1) (k_c + bk_\phi)^{\frac{1}{2n+1}}} \left(\frac{3W}{D^{1/2}}\right)^{\frac{2n+2}{2n+1}} \qquad (7)$$

where

n = exponent related to soil stratification

k_c = cohesive modulus of deformation

k_ϕ = frictional modulus of deformation

W = normal lunar weight

b = width of wheel

D = diameter of wheel.

Values for n, k_c and k_ϕ can be determined by special testing procedures (5), but on Earth in loose, dry sand, $k_c = 0$ and $n = 0.5$ to 1.0.

Figure 1 compares R_s with b, W, and D for a very weak but frictional dust ($n = 1.0$, $k_c = 0$, $k_\phi = 1.0$).

Example: let

W = 10,000 lb$_\oplus$ (1667 lb$_\leftmoon$)

D = 80 in.

b = 15 in.

Substituting in equation (7) gives R_s of 225 lb$_\leftmoon$, (or 41 times the hard surface resistance of the rigid wheel).

The mechanics of the interrelation between a pneumatic tire and soft ground are extremely complex. M. G. Bekker (5) gives a formula for the external, or soil, resistance.

Example: Let: W = 10,000 lb, D = 40 in., b = 10 in.
↑ Up D = 40 in. TO b = 10 in. Line
→ Across to Base Line
↗ Up Weight line to W = 10,000 lb ⊕
← Over to R_s = 430 lb ☽

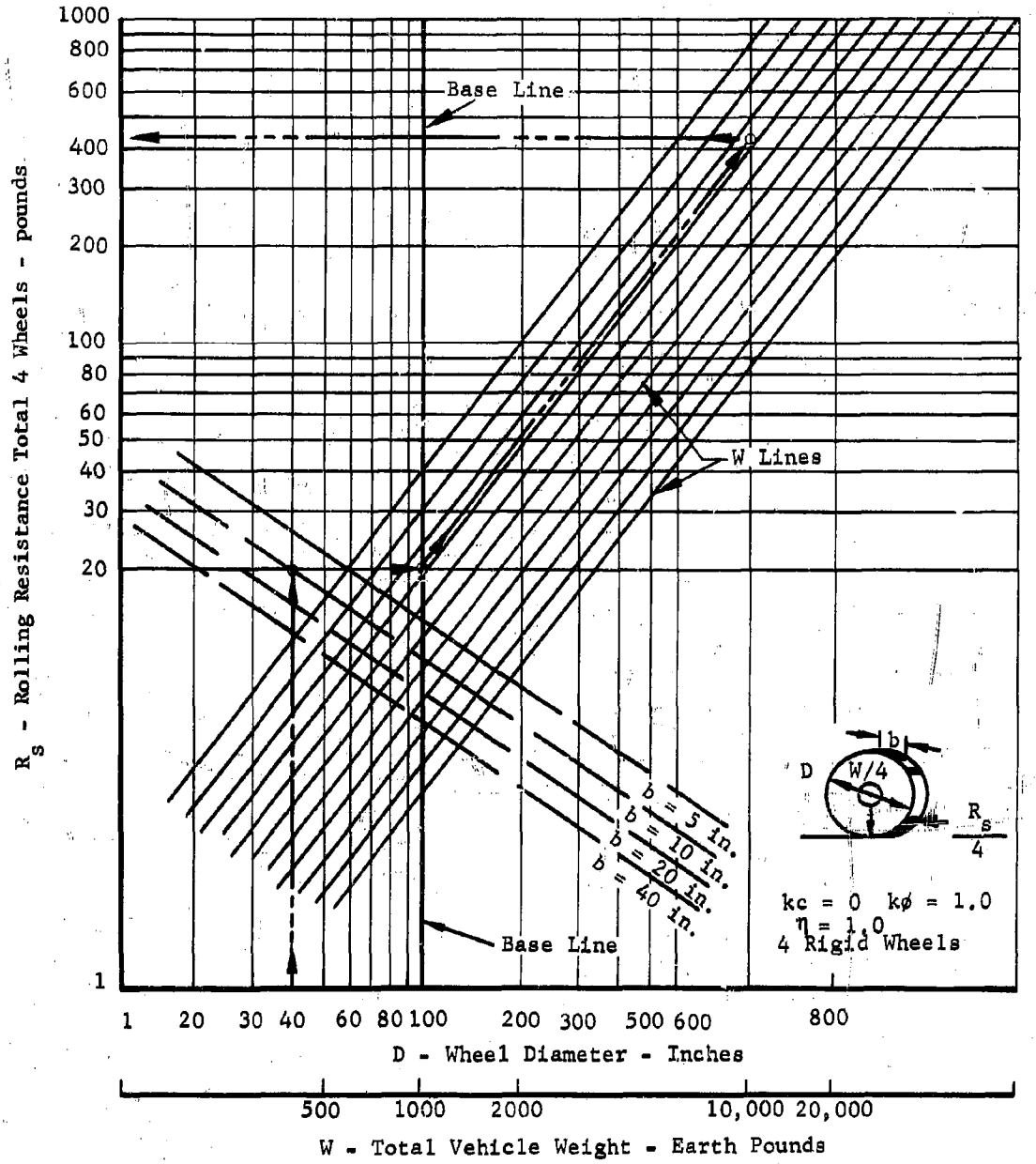


Fig. 1. Rolling Resistance vs D, b, W on Loose Frictional Lunar Soil

$$R_c = \frac{b\,(p_i + p_c)^{\frac{n+1}{n}}}{(k_c + bk_\phi)^{1/n}\,(n+1)} \qquad (8)$$

where

b = wheel contact width, (k_c, k_ϕ, and n as before)

p_i = unit pressure on the ground due to internal, or gas, pressure

p_c = unit pressure on the ground due to carcass stiffness.

To this must be added the internal resistance R_h to find the total rolling resistance of the tire.

Example: let

W = 10,000 lb $\oplus$ (1667 lb$_{\mathcal{C}}$)

b = 15 inches

D = 80 inches

$(p_i + p_c)$ = .75 psi.

This gives a static tire deflection of approximately 4.5 inches for an 80-inch diameter tire or a contact length of 37 inches.

Substituting these values in Equation (8) gives R_s of 16.8 lb$_{\mathcal{C}}$ for 4 wheels. $R_h \cong$ 33.3 lb$_{\mathcal{C}}$ (carcass resistance from Eq. 5) giving a total resistance of 50.1 lb$_{\mathcal{C}}$ (22 percent of rigid wheel resistance).

Note that in this example carcass flexure loss ($R_h \cong$ 33.3 lb$_{\mathcal{C}}$) is twice the amount of soil loss (R_s = 16.8 lb$_{\mathcal{C}}$). R_s is directly affected by internal pressure p_i which should be low on weak soil to keep R_s low. On the other hand, R_h is inversely affected by low internal pressures p_i. There is therefore an ideal internal pressure for every tire which produces the minimum rolling resistance for the particular load and soil involved (8). Other considerations which limit the softness of carcass p_c and pressure p_i are scuff resistance and excessive body movement. On the other hand, a tire may be too stiff for a specific soil condition. In this case it would act as a rigid wheel, and R_h would be low but R_s would be very high. To determine accurate values of rolling resistance and slippage it will be necessary to build and test proposed wheels on hard and soft surfaces under postulated environment to properly determine the carcass pressure and carcass resistance factors involved.

While this discussion refers to soft pneumatic wheels, the same factors can be applied to any flexible wheel. In the case of a metal-elastic wheel,

the internal pressure p_i would be zero, thus the carcass pressure p_c is the total ground pressure which effects the external rolling resistance. It may be possible to approach the low internal resistance of a rigid wheel with a flexible wheel made of an elastic material such as steel; however, the internal damping characteristic would approach zero and would be undesirable for a soft flexible wheel. Chassis instability, tread to surface chatter, or erratic tread performance may occur, particularly during acceleration, braking, and turning maneuvers. It would be safer at this time to assume that f_t will have approximately the same value for metal elastic wheels as for pneumatic tires.

To further complicate the problem, it must be noted that work along these lines has been based on towed wheels. Wheels under propelling torque undoubtedly will produce somewhat different results. In addition, bulldozing, soil drag, turning and scuffing forces add to the complexity of the problem.

One consideration which may ease the situation would be the ability of man to regulate the internal pressure of a pneumatic or flexible wheel. He could reduce the pressure on soft soil for maximum flotation and minimum resistance and increase the pressure on hard surfaces for improved efficiency, wear, maneuverability and stability.

The soft soil external resistance for a tracked vehicle is given (4) by

$$R_s = \frac{1}{(n+1)\,(k_c + bk_\phi)^{1/n}} \left(\frac{W}{s}\right)^{\frac{n+1}{n}} \tag{9}$$

where

s = length of track

b = width of track

with n, W, k_c and k_ϕ as before. Adding the internal losses R_h as found on hard, smooth surface gives the total rolling resistance for the tracked vehicle.

Example: let

W = 10,000 lb ⊕ (1667 lb ☾)

b = 15 in.

s = 50 in.

n = 1.0

k_ϕ = 1.0

k_c = 0.

(loose sand as before). Substituting these values in Eq. (9) gives R_s of 9.2 $lb_{\mathbb{C}}$ for 4 tracks per vehicle. To this we add R_h of 66.6 lb mechanical resistance Eq. (6) to arrive at R_{total} of 75.8 $lb_{\mathbb{C}}$, or 34 per cent of rigid wheel resistance.

Note that external resistance R_s of the track is only 12 per cent of the total, and that this value applies to an extremely weak soil. For longer tracks or slightly stronger soils, the external resistance can be neglected. It is thus evident that for tracked vehicles operating on frictional lunar surfaces, regardless of the soil strength, the rolling resistance is constant, determined for the most part by the mechanical, or internal, efficiency of the track.

Figure 2 compares the rolling resistances per pound for a rigid wheel, a flexible or pneumatic wheel, and a tracked vehicle, loaded with 250 $lb_{\oplus}$ as a function of wheel diameter or track length, while operating on the possible lunar surface extremes; a very weak frictional soil and a hard smooth surface. Figure 3 gives the same comparison for a 5000 $lb_{\oplus}$ load.

For the heavy load on assumed soft soil, the pneumatic wheel and track have equal resistances at 72-in. diameter and length. Slippage losses would be somewhat greater with the wheel; therefore, increasing the wheel to an 80-in. diameter should equate it to a 72-in. track. Increasing length of contact (D or S) beyond this point does not lower resistance significantly.

When operating on stronger soils, the flexible wheel is much more efficient than the track, particularly if wheel tread pressure is increased. As the weight of the vehicle is reduced, the advantage of the wheel becomes more apparent,

Figures 2 and 3 indicate that the rigid wheel cannot be made large enough without excessive size, weight and inertia to compete with a flexible wheel or track under the heavy load and assumed soft soil conditions; however, on a hard smooth surface the rigid wheel is the most efficient. With a lightweight vehicle operating on soft soil, the resistance advantage of a flexible tire or track is slight and a rigid or semi-rigid wheel might be favored for its simplicity and reliability.

Flotation on Soft Soil

If the vehicle does not deform the soil beyond its yield strength, it can be considered to float on the surface; in this case the resulting soil resistance is very low, approaching that of a hard surface. As can be seen in the previous analysis, the major part of the resistance of a rigid wheel would be eliminated, and the resistance of a flexible wheel would be reduced significantly but have negligible effect on the resistance of a track.

The maximum uniform static load W which a rigid surface may carry and maintain this flotation is given (5) by

$$W_{(max)} = 0.5 \gamma \quad b^2 s \, N_\gamma \quad \text{(dry frictional soil)} \quad (10)$$

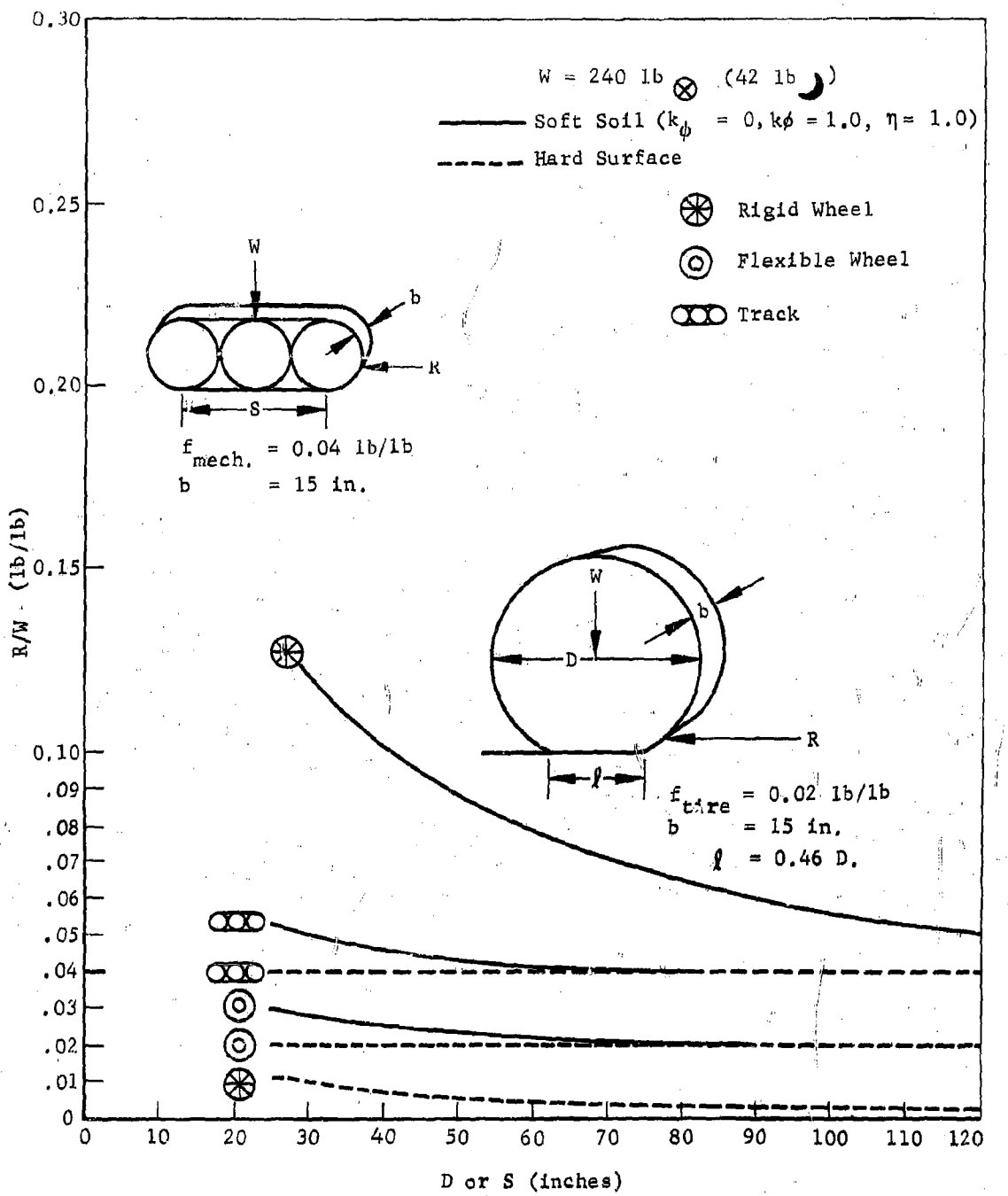


Fig. 2. Rolling Resistance

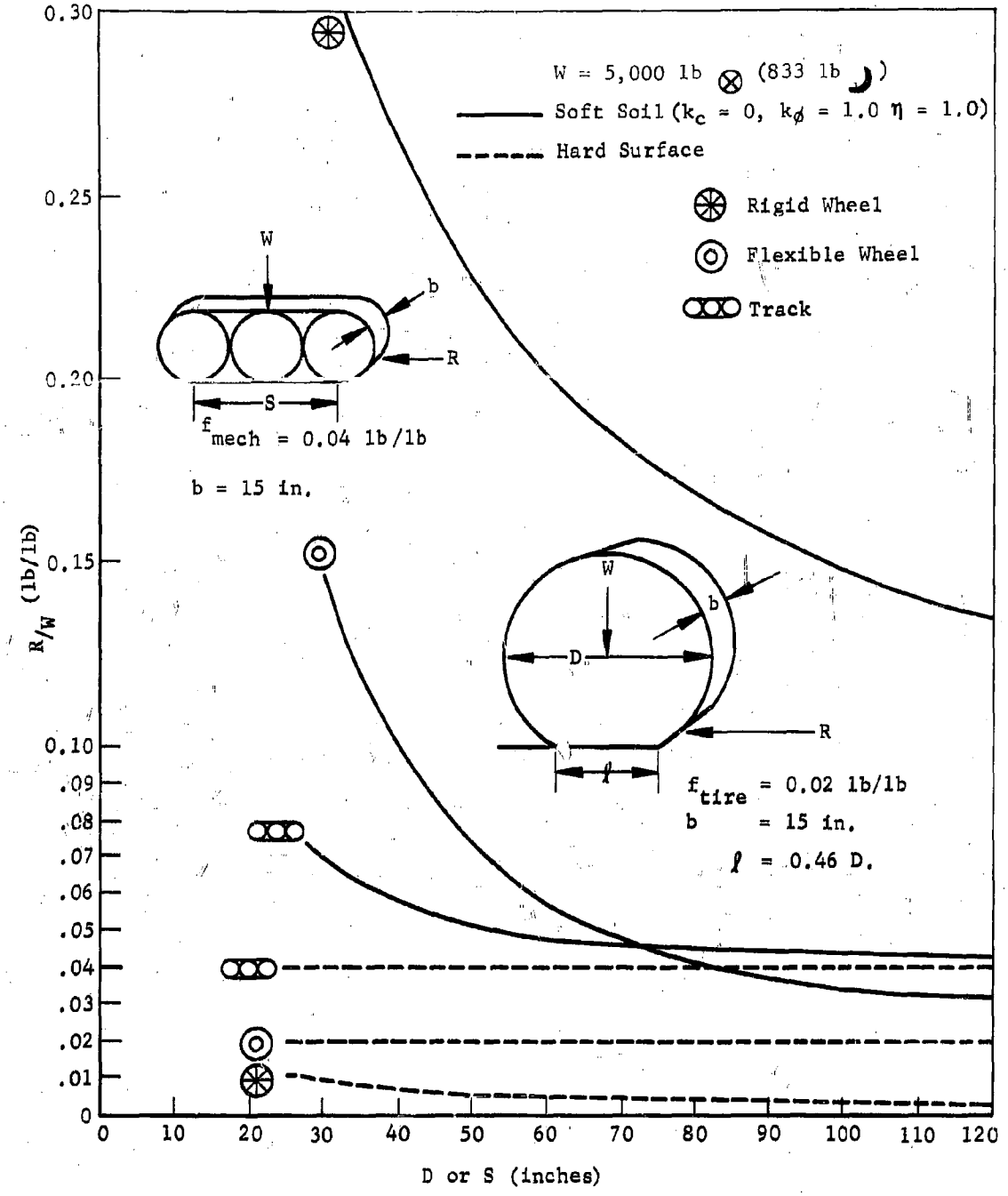


Fig. 3. Rolling Efficiency

where

γ = specific weight of soil

b = contact surface's smallest dimension (usually width of tread)

s = contact surface's longest dimension

N_γ = Terzaghi's coefficient for soil bearing capacity.

For the maximum weight a rigid wheel may carry on the same soil as before, let

b = 10 in. wheel contact length

s = 15 in. width

N_γ = 20 (if ϕ = 30 deg) (4)

$\gamma_{\leftmoon}$ = .008 lb/in.^{3} lunar sand at 50 percent voids

$\gamma_{\oplus}$ = .05 lb/in.^{3} Earthsand at 50 per cent voids.

Substituting in Eq. (10) gives W_{max} of 720 lb$_\oplus$ (120 lb$_{\leftmoon}$) for each wheel. Thus, the maximum flotation weight with 4 rigid wheels is 4800 lb$_\oplus$ (2880 lb$_{\leftmoon}$).

For a flexible wheel where deflection of the tire equals 11 percent of the diameter and assuming diam. = 80 in., s = 50 in. (contact length with 8.8 in. deflection), b = 15 in. width, $\gamma_{\leftmoon}$ = .008 lb/in.^{3} and N_γ = 20 as before, W_{max} = 5400 lb$_\oplus$ /wheel (900 lb$_{\leftmoon}$/wheel.) Thus for a 4-wheel vehicle, W_{total} = 21,000 lb$_\oplus$ (3500 lb$_{\leftmoon}$) which is the maximum static load and still maintain flotation.

A tracked vehicle with 4 tracks 50-in. long by 15 in. wide would have approximately the same flotation capacity of 21,000 lb$_\oplus$ (3500 lb$_{\leftmoon}$).

It must be remembered that these safe flotation loads are based on uniform static loads over the total contact surface. Concentrated or dynamic overloads during travel will be much greater and may cause an otherwise safe load to be temporarily exceeded. A pneumatic or flexible wheel provides better flotation under dynamic conditions, since the soil load is reduced by the increase in contact area and by the attenuation of the shock load resulting from the increased tire deflection.

The unknowns of lunar soil (γ and N_γ) and the flexibility of the wheel must be determined before more accurate flotation analysis can be completed; however, these examples indicate the ease with which flotation of vehicles with flexible wheels and tracks can be obtained on soft soils as compared to those with rigid wheels. If high values for γ and N_γ are found, a small size or medium-pressure flexible wheel should provide sufficient flotation for efficient operation of most vehicles.

Resistance Due to Surface Geometry

There are two classifications related to surface geometry: grade resistance and obstacle resistance.

$$\text{Grade Resistance, } R_g = W \sin \alpha \tag{11}$$

where W equals vertical weight and α equals grade angle. On a grade of 15 deg, for example, $R_g = .26\ W$. If all wheels are powered and a maximum coefficient of friction of 0.7 is assumed, the vehicle must use 38 per cent of its available surface thrust to overcome this grade resistance. Rigid wheel, pneumatic wheel or track vehicles must overcome equivalent grade resistances.

To arrive at the maximum slope the vehicle could climb, values for the resistance of soil, other obstacles, towed loads, etc. must be deducted. The momentum of the vehicle provides an added capability, and although the speeds are low (10 mph), the kinetic energy of the vehicle is equal to that on Earth, while the grade requirement, or change in potential energy, is but one sixth of that for Earth; hence the momentum effect should be helpful in climbing obstacles and short slopes.

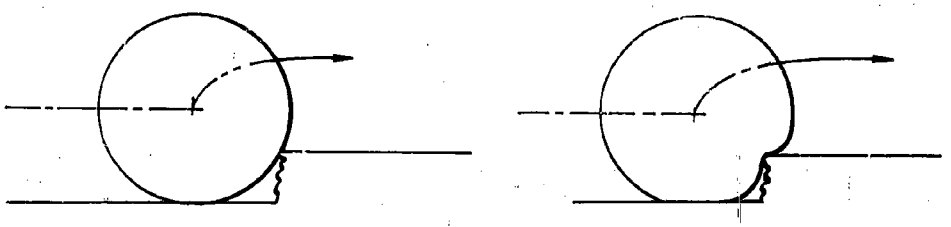
Obstacle Resistances (Depressions and Protrusions)

Accurate determination of obstacle performance is extremely complex and depends on surface factors, obstacle shape, vehicle configuration, speed and suspension characteristics (9). For this preliminary analysis, the vehicle was assumed to have 4 rigid, unsprung wheels of equal size with the center of gravity located near the vehicle center. An assumed quantity and size of obstacles and crevices was used in later energy-per-mile calculations.

The obstacles were assumed similar to a step in shape, where the maximum resistance at the start of climb is gradually reduced until it reaches zero as the centerline of wheel passes over crest of obstacle. Actual resistance will be less severe due to sloping obstacle sides plus the deflection of wheel or obstacle. Resistances would be greatly reduced for obstacles affecting single wheels, if the vehicle employs independent or sprung wheel suspension.

A soft, flexible tire has the capability to roll over many small obstacles with very low resistance by effectively swallowing the object or at least substantially reducing the maximum resistance through tire deflection (see Figure 4). In addition, the kinetic energy of the vehicle is dissipated more slowly upon impact due to greater wheel deflection. This effect reduces the shock load and may improve climbing ability if the wheel in contact with the obstacle is driven.

A tracked vehicle has a greater pitch stability problem due to teeter-totter effect as the center of gravity crosses the peak of the obstacle (see Figure 5), but the total obstacle resistance is approximately equal to that of a wheeled vehicle.



Rigid Wheel Flexible Wheel

Fig. 4. Rigid and Flexible Wheel Climbing Obstacle

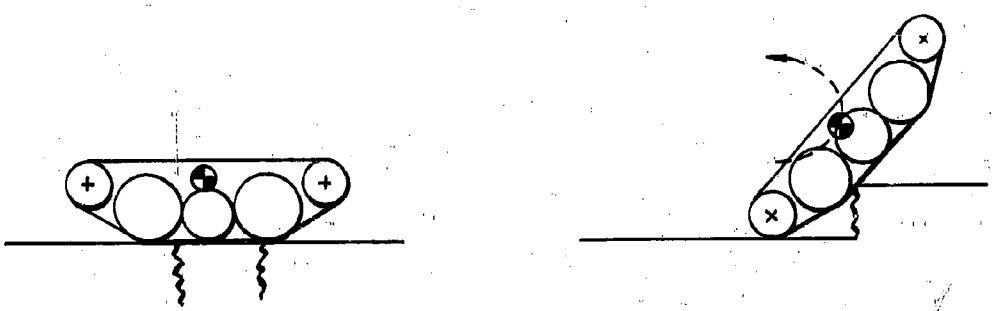
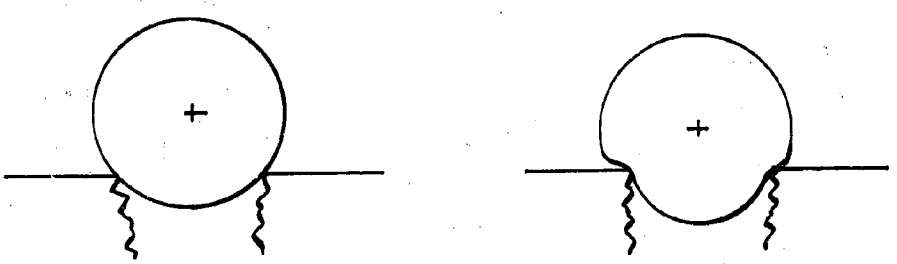


Fig. 5. Tracked Vehicle Crossing Crevice and Obstacle



Rigid Wheel Flexible Wheel

Fig. 6. Rigid and Flexible Wheel Crossing Crevice

A tracked vehicle is much superior for crossing narrow crevices or bridging rough surfaces (Figure 5) if the depression width does not exceed one-half the track length, and the center of gravity is near the center of the vehicle. Over this width, the vehicle will pitch end first into the crevice.

A rigid-wheeled vehicle (see Figure 6) may be expected to cross a crevice equal to approximately three-fourths of its wheel diameter, surface friction permitting. Ground clearance must be considered, however, to prevent under-body interference.

A flexible-wheeled vehicle crosses a crevice similar to a rigid wheel except static deflections are greater, and must be considered for safe width and ground clearance. Surface friction requirements, however, should be less critical than the case of the rigid wheel.

A manned vehicle will be able to maneuver around and through many obstacles that would stop a robot (10). If the vehicle is stranded, man will be able to take emergency action which may consist of calling for help or getting out and pushing or digging out the vehicle. A winch or tow cable should be a valuable tool for such operations.

Acceleration Resistances

The average force required to accelerate a body is equal to mass times acceleration. For vehicle designs, the relationship is

$$R_{acc} = \left(\frac{cW}{g}\right)a \tag{12}$$

where c is the mass factor which adds the inertia effect of rotating equipment to the translational mass. For many common vehicles, 1.2 is an average value for c (2).

The maximum acceleration occurs when maximum soil thrust H is used for acceleration $R_{acc} = H$. However, $H = W \tan \phi$ from Eq. (3), and substituting in Eq. (12) gives

$$a_{max} = \frac{g \tan \phi}{c} \tag{13}$$

for soft functional surfaces.

Thus, the maximum lunar acceleration, regardless of vehicle weight, will be approximately 3.1 feet/sec^2 in dry, loose sand, where $\tan \phi = 0.7$ and $c = 1.2$. Actually, in loose soils, not all the available thrust can be used to accelerate the vehicle. Some thrust must be used to overcome other exterior resistances such as soil, grade or obstacle; hence the maximum acceleration will be considerably less than 3.1 feet/sec^2. On hard surfaces, the coefficient of friction μ replaces $\tan \phi$ in Eq. (13) and thus may permit somewhat greater accelerations.

Starting Resistance

When a vehicle starts to move from a static condition, an instantaneous force must be added to other resistances. This starting resistance results from lubricant freezing, metal-to-metal contact in bearings, material deformation (such as tire set), tread to ground adhesion, etc. The harsh lunar environment may make this a serious consideration, particularly if vacuum welding takes place between contact surfaces.

LOCOMOTION PROFILE (CAPABILITY)

Figure 7 gives a graphic comparison of the lurain capability of typical vehicles as they negotiate various surfaces. The horizontal lines near the top of the chart indicate the maximum surface thrust per lb of weight that the vehicle can develop on postulated lunar soil. The vertical bars indicate the resistance per lb. of weight generated by various surface conditions. The vertical displacement between the thrust and the resistance lines (H/W-R/W) is an indication of the reserve thrust remaining for acceleration, starting forces, towing loads, slippage losses, etc.

Under these particular assumptions the rigid wheel has the least reserve thrust on difficult surfaces, but on hard or level surfaces appears to operate satisfactorily. On soft surfaces combined with obstacles or grades, the flexible wheel appears best with tracks a close second.

If assumptions could be reduced and more accurate data on environment and missions obtained, then charts such as these could be plotted to accurately show vehicle capability and help in the selection of locomotion modes and power plants.

ENERGY REQUIREMENTS

Energy per Lunar Mile

Since the requirements for vehicle range are unknown, the energy requirement must be derived on a unit distance basis -- i.e., the average energy required to drive the vehicle one typical lunar mile. Calculations which follow are used to show a method of energy per mile determination and provide a feel for the comparison of wheels and tracks over the same surface. As more factual data on lurain become available, these calculations can be updated and performance based on probability functions evaluated.

The first example is based on a 10,000 lb vehicle with 4 rigid wheels, 80-inch diameter by 15 in. wide.

$$\text{Energy per mile} = E_{\text{soil}} + E_{\text{geometry}} + E_{\text{acceleration}} \quad (14)$$

$$E_{\text{soil}} = E_{\text{hard}} + E_{\text{soft}} \quad (15)$$

$$E_{\text{geometry}} = E_{\text{grade}} + E_{\text{obstacles}} \quad (16)$$

$$E_{\text{obstacles}} = E_{\text{protrusions}} + E_{\text{crevices}} \quad (17)$$

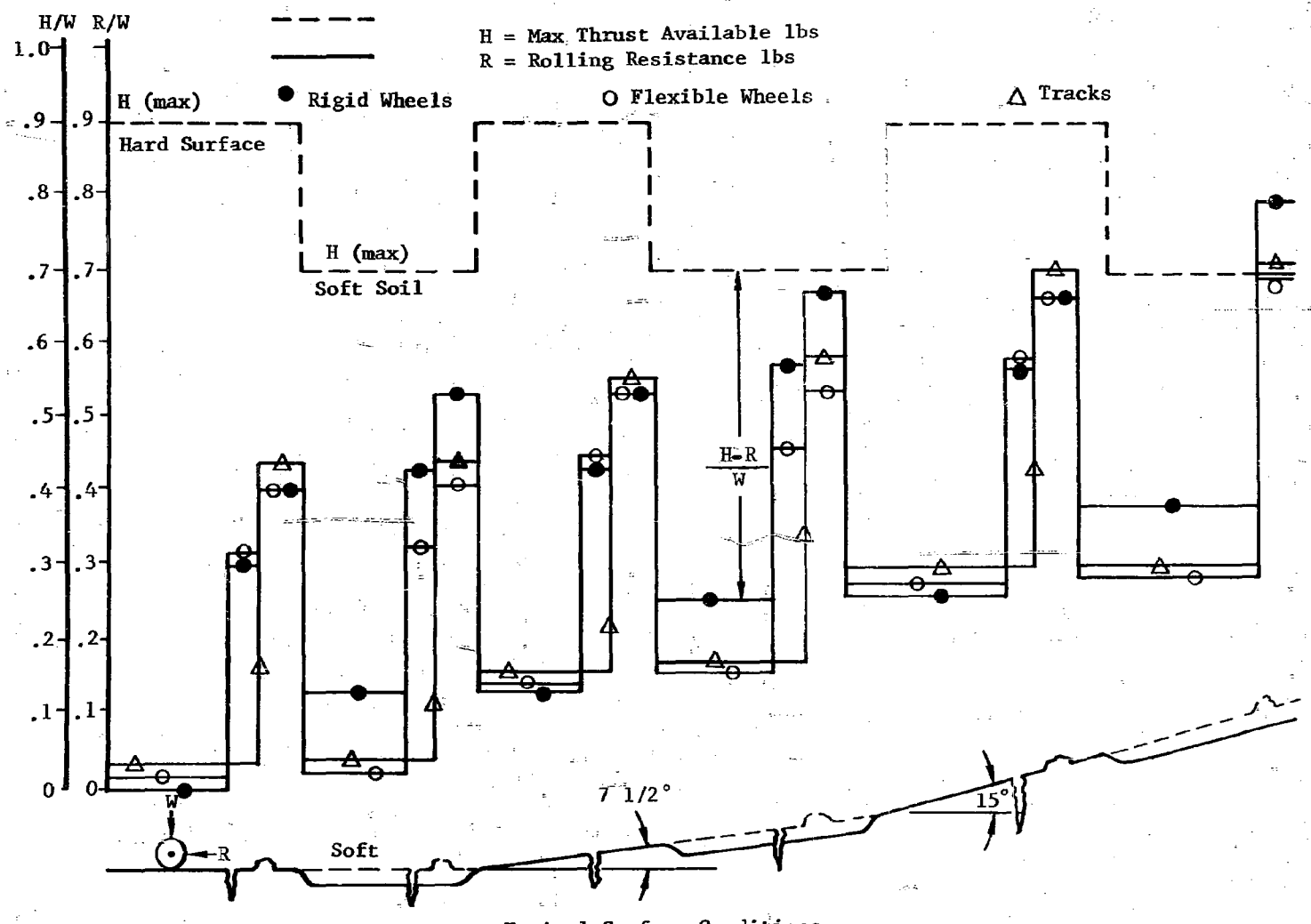


Fig. 7. An Example of a Locomotion Profile

Assume that one half of the distance traveled x_h is on hard surface, one half x_s on soft, and one fourth on x_g climbing grades averaging 7 1/2 degrees. In addition, assume one fifth x_a of travel consists of accelerating vehicle at an average acceleration of 1.1 ft/sec. Neglect the possibility of saving energy on down grades and during deceleration, thus adding to the margin of safety.

$$E_{hard} = R_h \times X_h \text{ where} \qquad R_h = 5.5 \text{ lbs Eq. (4)} \tag{18}$$

$$\therefore E_{hard} = 5.5 \times 0.50 \times 5280 = 14{,}500 \text{ ft-lbs/mile}$$

$$E_{soft} = R_s \times X_s \text{ where} \qquad R_s = 225 \text{ lbs Eq. (7)} \tag{19}$$

$$\therefore E_{soft} = 225 \times 0.50 \times 5280 = 595{,}000 \text{ ft-lbs/mile}$$

$$E_g = R_g \times X_g \text{ where} \qquad R_g = W \times \sin \alpha = 1667 \times .131 = 218 \text{ lbs Eq.(11)} \tag{20}$$

$$\therefore E_g = 218 \ .025 \ 5{,}280 = 288{,}000 \text{ ft-lbs/mile.}$$

$$E_{acc} = R_{acc} \times X_a \text{ (21) and } R_{acc} = 290 \text{ lbs } (a = 1.1 \text{ ft/sec}^2 \text{ Eq.(12)} \tag{21}$$

$$\therefore E_{acc} = 390 \times 0.20 \times 5280 = 412{,}000 \text{ ft-lbs/mile.}$$

Assume in one mile: 1-48 in., 10-24 in. crevices, also 1-30 in., 10-10 in., and 100-5 in. high protrusions are crossed by all four wheels.

$\therefore$ E obstacle $\cong$ 92,500 ft-lbs (calculated from equations in Ref. (7) (8)

$\therefore$ E total $\cong$ 1,309,000 ft-lbs/mile.

A similar procedure was applied to a soft pneumatic wheeled and a tracked vehicle traversing the same surfaces (see Figure 8 for comparative results).

It can be seen from these examples that the acceleration and grade losses are an important part of the total energy requirement but do not favor one locomotion device over another. Obstacle losses are probably the most difficult to determine and are therefore the most arbitrary; but from a comparative standpoint these examples indicate an insignificant difference. Flexible wheels would be more efficient for surfaces covered with small raised obstacles and tracks would be better for thickly creviced surfaces.

While Figure 8 indicates a considerable difference between locomotion devices on hard surfaces, the effect on the total energy requirement is small. However, on soft surfaces the greater resistance of rigid wheels and their effect on the total energy is evident. As vehicle weight is increased, these differences become even more pronounced.

Results cited neglect contact surface slip and mechanical losses in the drive system. The slippage losses vary with the locomotion mode (generally higher with rigid wheels) surface characteristics, and quantity of thrust

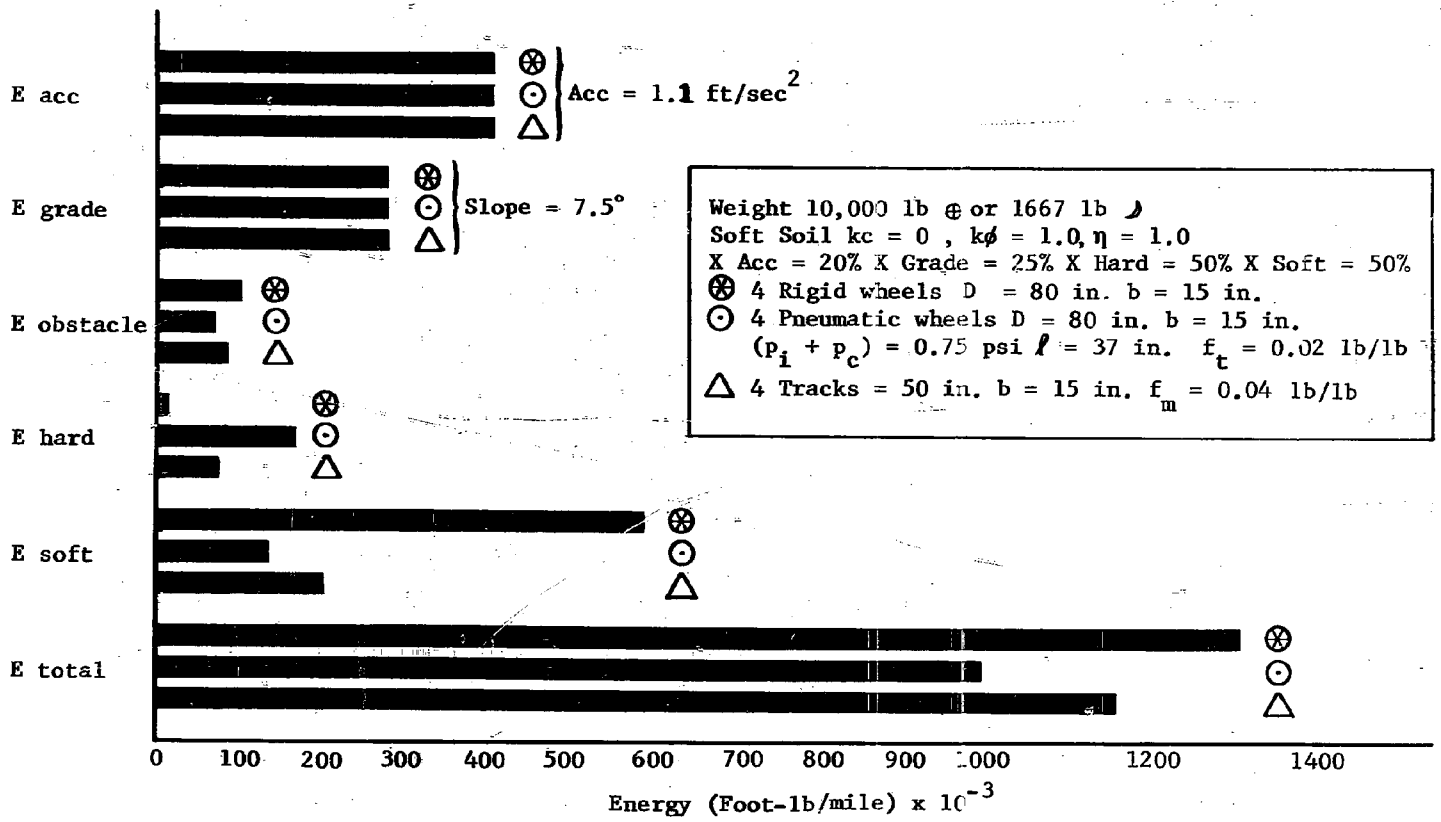


Fig. 8. Energy Requirements per Lunar Mile

required. For example, if the vehicle is attempting to climb a steep, soft grade, slippage would be a serious consideration. However, on firm surfaces with a high μ, or on soft but level surfaces, slippage is usually of a low magnitude. Turning losses due to tread slewing can be quite large on tracked vehicles but are generally low on wheeled vehicles unless the contact area is excessive, as, for example, on large, wide, low-pressure tires.

Comparative Range

The maximum range for a vehicle is a function of the total energy available for locomotion and the amount consumed per mile Eq. (14). Assuming a fixed quantity of fuel, the range can be estimated as follows:

$$\text{Range} = \frac{\text{Energy available} \times \text{efficiency of power system}}{E_{\text{total}} \left(1 + \text{misc. losses due to slippage, tread drag, etc.}\right)} \quad (22)$$

Figure 9 compares the range of a rigid and flexible wheel and a tracked vehicle. Values were taken from Figure 8 with revisions for 100 per cent hard and 100 per cent soft surfaces. Estimated miscellaneous locomotion losses of 25 per cent for rigid wheels, 15 per cent for flexible wheels and 10 per cent for tracks were added to the soft surface operations.

On 100 per cent soft surfaces the pneumatic tire has twice the range of rigid wheel and a 10 per cent greater range than track. However, on a 100 per cent hard surface the rigid wheel easily out-distances the others. If the internal losses of the flexible wheel could be reduced by increasing its tire pressure when on hard surfaces, it would equal the range of the rigid wheel as shown by the dotted line. Hence the pneumatic or flexible wheel could provide the greatest range under variable soil conditions.

DYNAMIC ANALYSIS

The dynamic analysis of surface vehicles is a complex subject worthy of many separate papers. This paper will therefore touch only upon some of the primary considerations.

It is well known that gravitational forces on the moon will permit man to jump great distances, and that the weight and static loads of an object on the moon are one-sixth of those on Earth. What seems to be often overlooked is that the mass of an object on the moon is equal to its mass on Earth. Therefore, since the force required to accelerate an object is equal to its mass times acceleration ($F = Ma$), the acceleration forces F on the moon are approximately equal to those on Earth. Dynamic loads may vary somewhat from this due to lunar environment, for example: lack of air damping, extreme temperature effects and the elastic effect of lunar soil compared to Earth soil. However, it should be noted that Earth-testing of the roving vehicle may not be as difficult as might be expected, since most of these dynamic loads are approximately equal.

It will be advisable to design lifting devices to accelerate and decelerate the load slowly to take full advantage of the moon's low gravity.

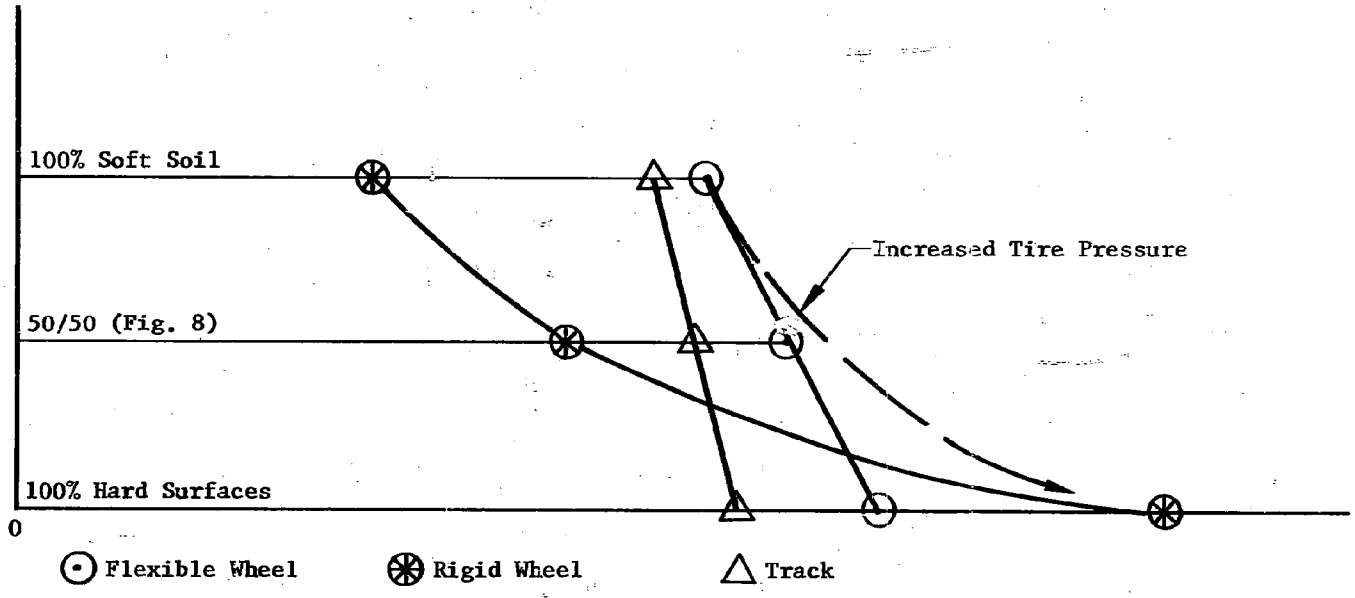


Fig. 9. Comparative Range

Ideally in a perfect system, the work done in lifting an object is equal to its weight times the distance and is independent of velocity and acceleration (neglecting air resistance). However, in actual practice, some energy is lost as internal friction and the acceleration forces may not balance out with the deceleration forces (during coast phase) due to control error. On Earth, these details are usually neglected, but on the moon, where weight and energy are limited, slow accelerations will be desirable.

Due to the lower gravitational force on the moon, free fall accelerations are equal to one-sixth those on Earth. Therefore, a body falling from the same height will have a vertical impact velocity approximately 2.5 times greater on Earth than on the moon. Drop tests on Earth will therefore be from one-sixth the height expected on the moon.

In crossing crevices and depressions at a constant speed on the lunar surface, the vehicle's wheels will not drop as far into the depression as on Earth. This again is due to the gravitational difference. The ride is therefore less severe than might be expected crossing these depressions (see Figure 10).

For a simple wheel crossing raised obstructions such as rocks, etc., and deflecting in a manner similar to an inclined ramp, the effects are noted on Figure 11. Actual performance will differ somewhat from these examples depending on bump shape, speed, and elasticity of wheel and obstacle, along with the suspension characteristics of the vehicle.

The overturn (rotational) effect of a ramp-type bump also points up an interesting consideration: In order for a vehicle to overturn, the center of gravity must be rotated until it is directly over or past the rotational center (r.c.). (See Figure 12).

At this point, the vehicle has changed its potential energy by Δ P.E. = (W · h). If the kinetic energy of rotation put into the vehicle by the bump is greater than this change in potential energy, the vehicle will overturn or

$$\text{K.E.} = I_m \frac{\omega^2}{2} \text{ if } \geq W\,h \text{ (vehicle overturns)} \tag{23}$$

$$I_{m_{\leftmoon}} = I_{m_{\oplus}}, \quad W_{\leftmoon} = \frac{W_{\oplus}}{6} \text{ and } \omega = (\tan \beta)\,\frac{V}{\ell}\,. \tag{24}$$

The restorative component (Wh) due to lunar weight is 1/6 of that on Earth, but the overturning component (KE), a function of the rotational moment of inertia I_m and velocity (V), remains the same; therefore, the vehicle will overturn much easier on the moon. It should also be noted that the danger of overturn increases as the square of the vehicle speed.

The example given is simplified to point out the problem that exists. The danger of overturning must be analyzed with respect to the effects of different vehicle and suspension configurations, gyroscopic effect of wheels, and

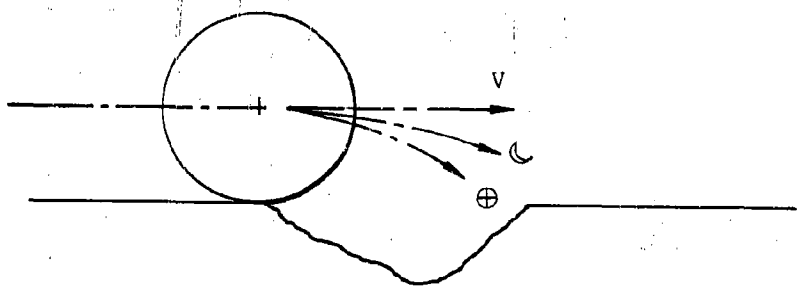
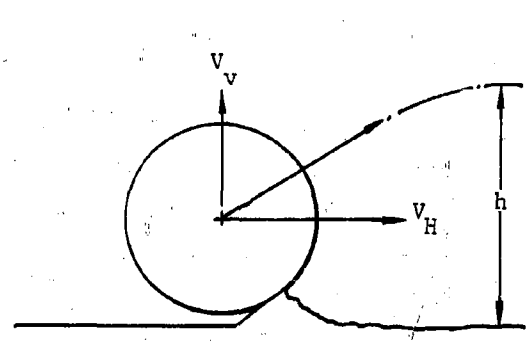


Fig. 10. Depression Crossing (Free Body)



$V_{v_{\smile}} = V_{v_{\oplus}}$ (Velocity leaving top of bump)

$h_{\smile} = 6h_{\oplus}$ (Height of wheel trajectory, assuming free body)

$V_{v_{\smile}} = V_{v_{\oplus}}$ (Vertical velocity at return impact with the ground)

Fig. 11. Raised Obstacle Response (Free Body)

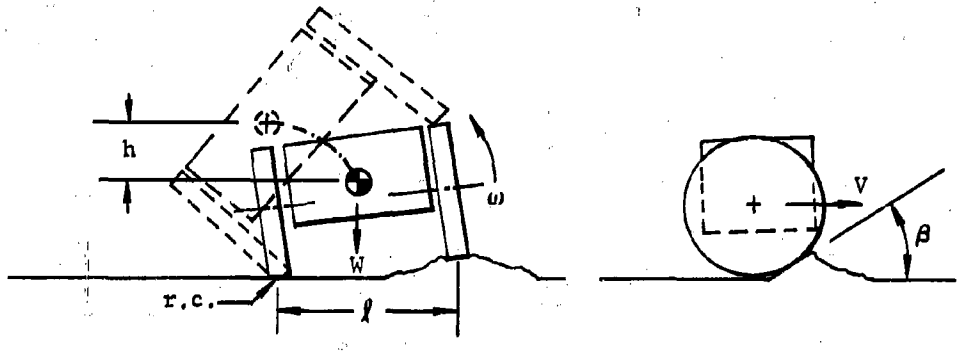


Fig. 12. Lateral (roll) Stability

variations in obstacle characteristics. It appears that slow speeds, a wide wheel stance, and a low center of gravity will be desirable for surface vehicles, especially if the surface is rough or has many raised obstacles.

If vehicle speeds are low enough, it may not be necessary to add the weight and complexity of a spring suspension. At low speeds, even with extremely soft suspension systems, many reoccurring bumps may generate forcing frequencies at or near the vehicle's natural frequency. In these cases, the spring suspension may actually amplify the bump effect. As vehicle speeds are increased, the probability of bumps generating resonant conditions is reduced and the attenuating effect of suspension increases. A soft suspension, which would be required to give worthwhile shock protection at slow speeds involved, would require a good damping device to prevent this resonance and limit the large body movement related to such soft suspensions. Also, with a sprung suspension system and the low gravity effect, wheel hop becomes critical and must be considered.

The possibility of using a soft flexible wheel with semi-rigid body mounting rather than a spring suspension system should be investigated. The tire pressure could be varied to provide suspension characteristics best suited to the surface conditions and vehicle speeds involved.

BRAKING ANALYSIS

If the vehicle is operated by a man on or near it, the communication time delay problems associated with Earth-to-Moon remote control are eliminated (11). While this eases the reaction control problems, the braking distances required to stop the vehicle are still approximately six times as great as on Earth due to the gravitational difference. The kinetic energy to be overcome is the same, but the maximum braking force, which is proportional to the normal weight, is 1/6 of that on Earth.

As with traction, the braking force depends on the surface (soil), and tread characteristics. However, making the conservative assumption that the wheel is sliding and the ground is hard and smooth, permits a simplification of the braking analysis. Thus, vehicle weight is cancelled out and the braking help from soil, such as compaction, bulldozing, and drag may be neglected. Therefore, the minimum stopping distance for all-wheel braking on level ground varies as the square of velocity and inversely as the coefficient of friction.

$$\text{Stopping Distance} = \frac{V^2}{2g\mu} \tag{25}$$

For example, at five miles per hour and μ of 0.75, stopping distance is 6.8 feet$\leftarrow$ (1.13 $\oplus$ feet). While at ten miles per hour, distance increases to 27 feet$\leftarrow$ (4.5 feet $\oplus$). If the coefficient of friction drops to 0.50, the distance at 10 mph increases to 41 feet$\leftarrow$ (6.7 feet $\oplus$).

Braking on a down slope increases the stopping distance depending on the grade angle. For example, at 10 miles per hour, a coefficient of friction of 0.75 and a grade of 15 degrees, the braking distance is approximately 41 feet$\leftarrow$

for all wheels sliding. On slippery surfaces (low μ) the slope angle is very important. For example, if the coefficient of friction drops to 0.25 or less, the vehicle cannot be stopped on a 15 degree or steeper slope regardless of its speed.

A comparison of the fundamental mechanics of stopping the vehicle versus turning it to avoid an obstacle indicates that braking to a stop is safer, with the braking distance approximately one-half that required for turning.

CONCLUSIONS AND RECOMMENDATIONS

It is generally expected that the first operations will be performed on lunar areas where simple forms of locomotion such as wheels or tracks may be employed. As the exploration requirements become more extensive, more exotic vehicles, if required, may be attempted.

If the surface is weak and loose, rigid wheels will be limited to light loads or short range vehicles. This conclusion is due for the most part to the greater tendency of rigid wheels to sink and slip on weak surfaces. If reasonable soil densities (compressive strength) or thin layers of dust over strong subsurfaces prevail, the rigid wheel or smaller semi-flexible wheel will be preferred for many applications for reasons of efficiency, smaller mass and reliability (12).

A pneumatic wheel for daytime operation or a simple metallic flexible wheel would be extremely desirable for heavy manned vehicles if they can be developed to withstand the lunar environment. With the added ability to vary ground pressure to suit the surface and load conditions, this mode would provide a good ride and high efficiency on both strong or weak soils as well as on smooth or rough surfaces. Man should be able to perform the necessary maintenance and avoid most hazards which might damage this wheel. Selection of wheel size and material may be made when more factual data are available. It will be influenced by soil values, obstacle capability, ground clearance, vehicle stability and height, weight, inertia and suspension requirements.

A tracked vehicle offers advantages for specialized applications such as dozing, scooping, towing, and carrying topheavy equipment (13), due to its lower vehicle height, good stability, steering simplicity and reduced slippage in soft soils. The added weight, complexity and mechanical losses, however, may prevent its universal use unless the lunar surface is densely pitted or fractured (14).

The low gravitational force on the moon greatly reduces surface traction, vehicle accelerations, braking capabilities, energy and power requirements, static deflections and impairs dynamic stability. Other dynamic characteristics such as impact accelerations, and the resulting loads and material deflections, will remain equivalent to those on Earth since the mass of the vehicle remains the same.

These considerations combined with safety requirements and fuel limitations impose some unique design problems. First generation vehicles will

probably be designed for and restricted to slow speed operation. Flimsy construction so often depicted will not be practical. Independent drive systems will be employed at each wheel or track and possibly some sort of emergency or secondary locomotion device incorporated to improve mission reliability. Vehicles should be designed to withstand overturning or crash impacts during landing, unloading, and travel, and should be capable of being righted, with auxiliary equipment available to take over at subsequent failure of critical components.

The semi-rigid wheel appears practical for strong surfaces and for lightweight, short-range vehicles on weak soils. For heavy vehicles on weak surfaces, the choice between wheel and track is less pronounced, and while favoring the large flexible wheel could swing to the track for specialized applications or for unusual surface conditions.

REFERENCES

1. Phyllis Buwalda, "Lunar Environment, "Technical Release 34-159, JPL, October 1960.

2. J.J. Taborek, "Mechanics of Vehicles, Machine Design, May 30, 1957, June 13, 1957, June 27, 1957, July 11, 1957, July 18, 1957, July 25, 1957, August 8, 1957, August 22, 1957, August 29, 1957, September 5, 1957, September 19, 1957, October 17, 1957, November 14, 1957, November 28, 1957, and December 12, 1957.

3. G.K. Wehner, "Sputtering Effects on the Moon's Surface," ARS Journal, March 1961.

4. M.G. Bekker, "Theory of Land Locomotion - The Mechanics of Vehicle Mobility," Ann Arbor, Michigan: University of Michigan Press, 1956.

5. M.G. Bekker, "Off the Road Locomotion," Ann Arbor, Michigan: University of Michigan, 1960.

6. Z. Janosi and B. Hanamoto, "The Analytical Determination of Drawbar Pull as a Function of Slip for Tracked Vehicles in Deformable Soils," 1st International Conference on the Mechanics of Soil-Vehicle Systems. Torino, Italy, 1961.

7. "Functional and Dynamic Tests of Hawk Loader," NORAIR Report NOR-61-245.

8. Z. Janosi, "An Analysis of Pneumatic Tire Performances on Deformable Soils," 1st International Conference on the Mechanics of Soil-Vehicle Systems, Torino, Italy, 1961.

9. G.P. Rettig, and M.G. Bekker, "Obstacle Performance of Wheeled Vehicles," Land Locomotion Research Branch, Research and Development Division, O.T.A.C., March 1958.

10. W.B. Sponsler, "Preliminary Design Analysis of a Remote Controlled Lunar Roving Vehicle," Northrop Corporation Report N-269-61-9, June 1961.

11. James L. Adams, "An Investigation of the Effects of the Time Lag Due to Long Transmission Distance Upon Remote Control." Phase II - Vehicle Experiments. Phase III - Conclusions.

12. "A Feasibility Study and Conceptual Designs for Operational Unmanned Lunar Mission for 1964-65," Northrop Corporation Report, N-269-61-12, August 1961.

13. "Proposal to Establish a Design Concept for a Lunar Roving Vehicle," Northrop Corporation Report NB 61-345, November 1961.

14. John W. Salisbury and Charles F. Campen, Jr., "Location of a Lunar Base," G.R.D. Research Notes #70, Bedford, Mass.: Cambridge Research Lab, USAF, October 1961.

PENETRATION STUDIES OF SIMULATED LUNAR DUST

R. D. Rowe
E. T. Selig

Armour Research Foundation
Soil Mechanics Section
Chicago 16, Illinois

ABSTRACT

Results are presented for the static and dynamic penetration resistance of a simulated lunar dust in a hard vacuum environment. Specimens of finely ground silica, covering a range of densities, were tested at a number of absolute pressures from one atmosphere down to 5×10^{-8} torr. While the nature of their behavior was somewhat different, both static and dynamic penetration resistance were found to depend significantly on initial specimen density and on vacuum levels, increasing with an increase in density or a decrease in pressure.

The work described in this paper was Armour Research Foundation sponsored.

PENETRATION STUDIES OF SIMULATED LUNAR DUST

R. D. Rowe and E. T. Selig
Armour Research Foundation
Soil Mechanics Section
Chicago 16, Illinois

INTRODUCTION

In order to assure the success of lunar missions planned by this country for the next decade, scientists and engineers will need to acquire considerable knowledge regarding the nature of the lunar terrain, and the expected behavior of the surface material under environmental conditions corresponding to those existing on the moon. Systems must be designed, first for landing on the moon; then for performing such functions as material sampling and analysis and traversing of the surface. Later, remote controlled take-offs will occur and then, with the arrival of man, there will be the more complex tasks of exploration of the lunar topography and the construction of habitable protective shelters. The accomplishment of all of these tasks involves a knowledge of the mechanics of lunar "soils".

Most of the information assembled to date on the surface conditions has been obtained from telescopic observations employing the visible, infrared and radio wavelengths of the electromagnetic spectrum. A vast number of papers have been published in the technical literature covering these investigations. Available information permits certain conclusions and hypotheses to be drawn regarding the probable surface conditions on the moon.*

Lunar environmental conditions are characterized by the absence of an atmosphere -- surface pressure is estimated to range between 10^{-12} to 10^{-14} torr.** The gravitational field has an intensity of only one-sixth of that of the earth and the temperature ranges from -244° F to +273° F at the subsolar point. It is thought that only the lunar highlands and rims of craters are free of dust; in these regions bare rock may be encountered. In other areas the dust layer may vary in thickness from a few centimeters in open maria to hundreds of feet over the floor of large craters. The largest portion of the particle spectrum is expected to be in the 1 to 100 micron range, and composed of at least 40% silicon dioxide.

---

* "Probable Soil Conditions on the Moon and Terrestrial Planets", N. A. Weil, Proceedings, 1st International Conference on Mechanics of Soil-Vehicle Systems, Turin, Italy, June, 1961.

** Torr designates pressures expressed in millimeters of mercury.

The penetration tests described in this paper were the first phase of a study to determine the characteristics of the hypothesized dust condition which might exist on the moon. Basically two types of tests were considered: (1) dynamic penetration of rod-shaped projectiles dropped into the dust, and (2) static bearing tests with small blocks resting on the surface of the dust. These tests were chosen as a means of comparing the strength characteristics of the dust under various levels of vacuum, ranging from 1 atm. down to 5×10^{-8} torr. No attempt was made to simulate the differences in the temperature and gravity between the earth and the moon.

DESCRIPTION OF EXPERIMENTS

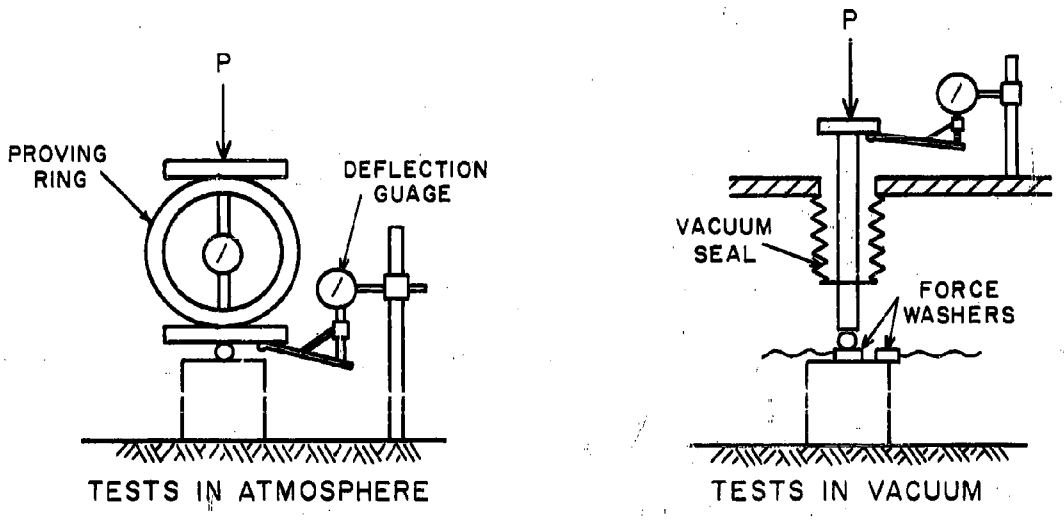
Materials and Specimen Preparation

The material selected to represent lunar dust was a white silica flour consisting of greater than 99 per cent silicon dioxide. It was prepared by finely grinding washed silica sand to provide a particle size distribution in which 85% was within the 2 to 40 micron diameter range. This material has a specific gravity of 2.75 and may be assembled with a bulk density ranging from a low of about 50 pcf (lb/ft^3)(voids ratio e=2.45) to a high of 85 pcf (e=1.02). At a density of 50 pcf the specimen is in a relatively unstable state, and compacts readily at the slightest jarring or perturbance. When compacted to 85 pcf the specimen is relatively hard and has a high strength.

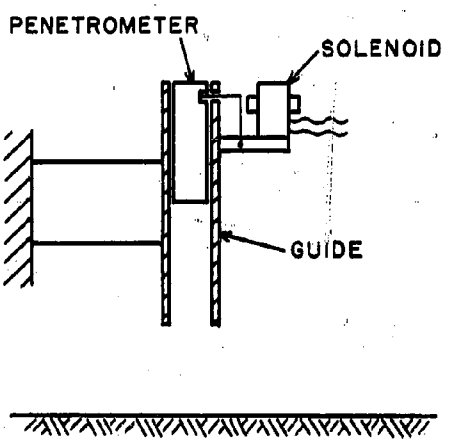
Low density (50-60 pcf range) specimens were prepared by a careful manual pouring of the dust from a fixed height. For preliminary tests at densities between 60 and 70 pcf the container was vibrated for a predetermined time for each layer of dust added. However, compaction under a slowly applied uniform surface pressure was found to be more effective and resulted in a greater specimen uniformity. This was accomplished with an inflatable rubber bag made to fit the cross-section of the container. A porous filter paper was placed between the specimen and the bag to allow the pore air to escape from the dust as the volume of the voids was decreased. Both the pressure and the number of layers were varied to alter the density. This method was subsequently used for obtaining densities of 60 pcf and higher.

The presence of moisture in the specimen causes considerable outgassing during evacuation. To reduce the draw-down time at the higher vacuum levels, the prepared dust bed was baked in an oven at 110°C for a minimum of 18 hours. The specimen was then transferred directly to the vacuum chamber (Fig. 1) and the evacuation process begun without utilizing chamber bake-out. In an attempt to obtain a higher vacuum level, for the last two penetration tests a bake-out temperature of 100°C was also maintained during the initial stages of the draw-down process.

For consistency, the specimens for the atmospheric tests were also oven dried at first. Even after several hours of reexposure to the air, there was no measurable moisture content in the oven-dried specimens, due to the fact that hygroscopic moisture content of air-exposed silica flour is only about 0.05 per cent. The drying procedure was later omitted from the atmospheric tests since it appeared to have no affect on the results.



a. STATIC



b. DYNAMIC

Fig. 2. Schematic of Penetration Apparatus

fore-line pressure to the diffusion pumps. This system without baking out, has reached a pressure of 2.6×10^{-9} torr in a period of four hours.

The experimental set-up shown in Fig. 1 consists of a rectangular pan 16 in. wide by 20 in. long supported in a frame to which the dynamic penetrometers are attached. It was believed that this size container provided sufficient surface area so that the 2-in. footing and four dynamic penetrometers could be tested in the same specimen without influencing each other. The piston for loading the footing passed through the top of the chamber and was sealed with a steel bellows permitting vertical motion. The dust was placed in this container to a depth of 5-3/8 in.

After the apparatus had been set up, one rod was released to determine the dynamic penetration of the specimen at one atmosphere. Additional rods were then released at several stages of successively higher vacuums as the draw-down progressed. At the final vacuum level the static penetration was conducted and the final rod released. The rod penetrations were all measured at the completion of the run after atmospheric pressure had been restored.

The draw-down process had to be sufficiently gentle to avoid disturbing the state of density of the dust specimen. Initial draw-down had to proceed at a rate slow enough to allow the air entrapped in the voids to escape without creating appreciable uplift pressure on the dust particles. When the flow was too rapid the material would separate in layers and rise in the container. Porous stones were placed in the bottom of the container to allow the pore air to escape from the bottom of the specimen as well as the top, permitting faster draw-down. Releasing of the vacuum at the end of the test also had to proceed slowly to avoid consolidation of the specimen.

Atmospheric Tests

A series of static penetration tests with the 2-in. square footing were conducted in the atmosphere to investigate the effect of density on specimen behavior and to provide a comparison with the same test under high vacua. Initially, to reduce specimen preparation time, smaller containers 13.5 in. in diameter were used and the dust placed to depths of either 5-3/8 in. or 3-5/8 in. In order to determine if the smaller specimen size had a significant effect on the static penetration results, an additional series of static penetration tests in atmosphere were performed using the same container and depth of layer as for the experiments in the vacuum chamber.

Low Vacuum Tests

To distinguish the effect of pore air pressures developed in the specimen under rapid loading from the effect of high vacua, dynamic penetration tests were conducted at vacuum levels in the range of 0.1 to 760 torr (atmospheric pressure). The high vacuum chamber could not be conveniently used for these low vacua so a bell jar was provided together with a manometer and a low-capacity vacuum pump (Fig. 3). A smaller specimen container was used and fitted with a single penetrometer identical to that for the high vacuum chamber. However, due to the more limited space in the bell jar, the drop height had to be decreased slightly.

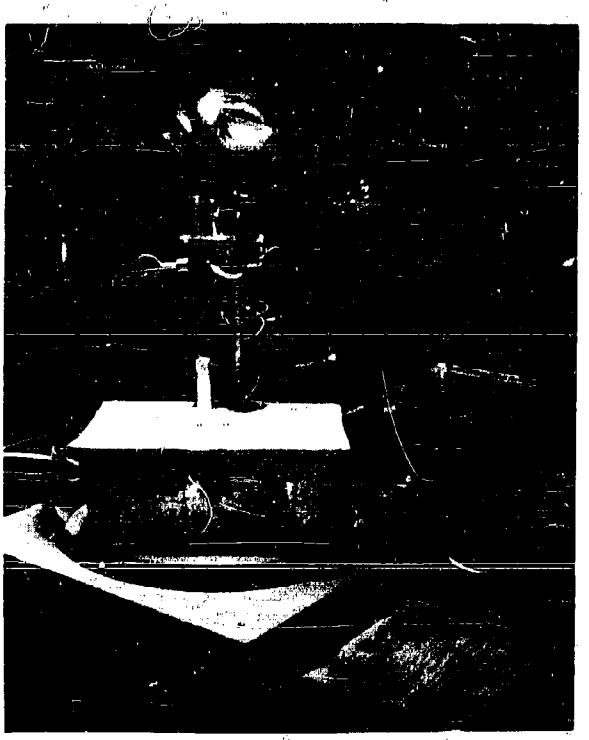


Fig. 3. Bell Jar Apparatus

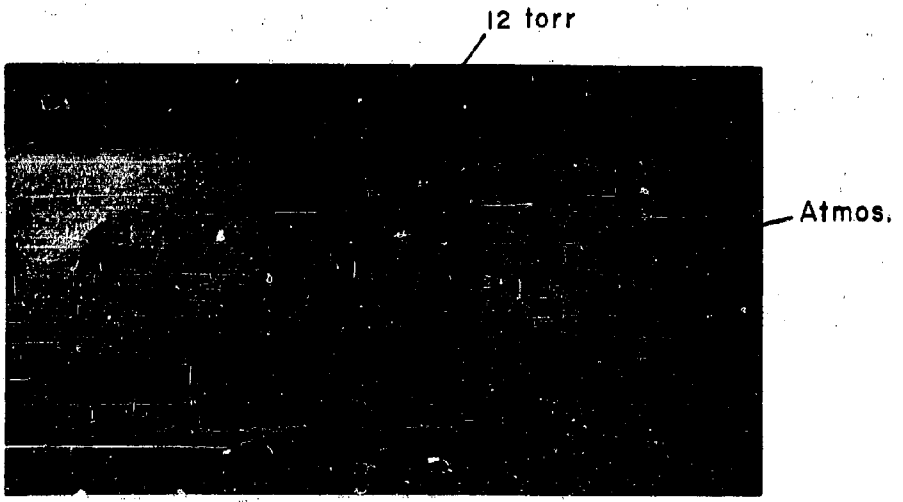


Fig. 4. Dynamic Crater Formation

STATIC PENETRATION TESTS

For a particular dry, cohesionless material and a specific size and shape of static penetrometer there are two factors which exert a significant influence on the penetration resistance: the bulk density of the soil and the absolute pore air pressure inside the voids. As the number of particles in a unit volume is increased the resistance to further consolidation and to internal shearing increases. This behavior is largely governed by interparticle friction which is expected to increase as thickness of gas layers adsorbed on the surface of the particles decreases and is finally disrupted. Thus, application of a sufficiently high vacuum should also affect penetration resistance.

Effect of Density Change

To establish the relationship between static penetration and density of the silica powder a series of tests were performed in atmosphere with densities ranging from 55 pcf to 80 pcf. Representative load-penetration curves obtained are shown in Fig. 5. At the lower densities the slope of the curves increased with penetration. This indicates a consolidation of the material directly beneath the footing. At the higher densities the slope decreased with penetration. As the density increases, the amount of penetration accompanied by consolidation is less and shearing failure of the surrounding material begins to take place.

Even at the highest densities achieved there is no evidence of a maximum bearing capacity. This is characteristic of local shear failure of soil associated with footings as described by Terzaghi.* Since no maximum load exists, the influence of density on penetration resistance can best be observed from a correlation of load with density for specific amounts of penetration. This correlation for 1 in. penetration is shown in Fig. 6. The selection of 1 in. for comparison was arbitrary -- an examination of the data for other penetration levels showed the same trends. The increase in load with density is quite pronounced -- for example, the load corresponding to 1 in. penetration at 78 pcf is about 19 times greater than that for the same penetration at 59 pcf.

Three sets of data are included in Fig. 6: results obtained in the small container for two layer depths and results from the large container at the same depth used for the vacuum experiments. There appears to be no significant difference in the results obtained under these three sets of conditions. The results in the large container are the most consistent and have been fitted with a smooth curve. The variation of the values as a whole is most likely due to the fact that the densities used for correlation were the average values for the entire specimen. Although, with the accuracy of instruments available, it was not possible to detect any variation of density within the specimen, it is assumed that a density variation of several per cent may have been present in the sample. This fluctuation, although small, can account for most of the data scatter because of the critical influence of density.

---

* Terzaghi, K., Theoretical Soil Mechanics, John Wiley and Sons, Inc. New York, N. Y., 1959.

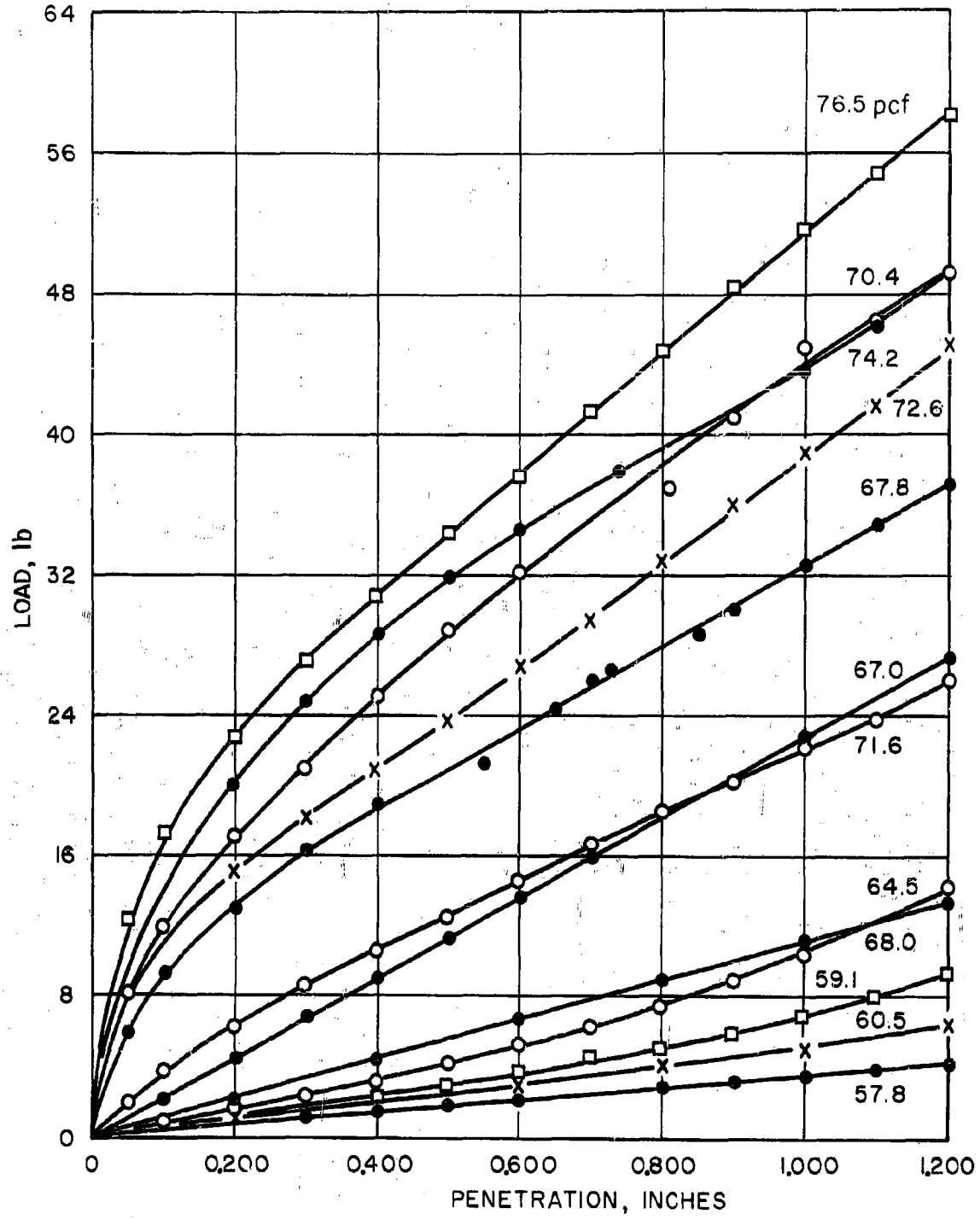


Fig. 5. Static Penetration Results in Atmosphere

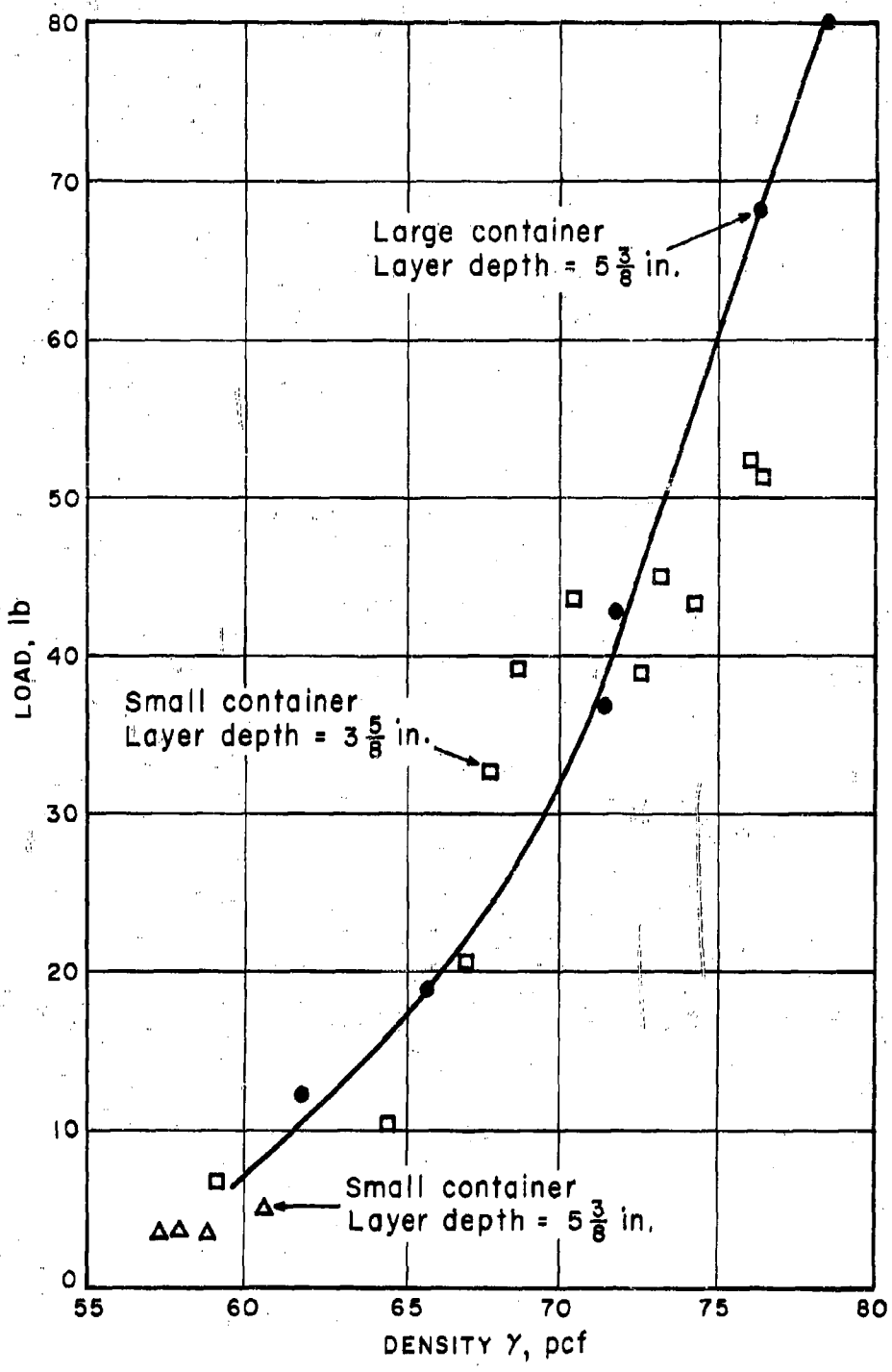


Fig. 6. Correlation of Penetration Resistance with Specimen Density

Effect of Absolute Pressure

Static penetration tests were conducted in the high vacuum chamber at various vacuum levels for densities in the vicinity of 65, 70 and 75 pcf. Typical results are presented in Fig. 7. The shapes of the penetration curves in vacuum are essentially the same as those in atmosphere.

To show the effect of vacuum level on penetration resistance, the loads at 1 in. penetration for densities of 65, 70 and 75 pcf were correlated with absolute pressure (Fig. 8). It was not possible to reproduce the specimen density exactly; thus, to minimize the influence of density variation in evaluating the effect of vacuum, the experimental curves in each density group (65, 70 and 75 pcf) were adjusted to the common density. The adjustments were made utilizing the curve in Fig. 6, i. e., by multiplying the load at 1 in. penetration by the ratio of the load at the common density to the load at that same density in atmosphere, for the same penetration. It is believed that this type of adjustment is reasonable for the small variation in density involved.

Figure 8 shows that up to at least 10^{-5} torr vacuum there is no significant change in static penetration resistance compared to that in atmosphere. Beyond this range the data for 65 and 75 pcf show a sharp increase; the data at 70 pcf show a steady, but not sharp, increase. All of these data, further adjusted to a common density of 70 pcf, are represented by the dashed curve in Fig. 8. While the qualitative validity of this latter adjustment may be questioned because of the wide range of density involved, it nevertheless serves to show the trend indicated by the data as a whole. While the penetration resistance unquestionably increases with vacuum in the 10^{-7} torr range, whether this increase is sharp or more gradual is not clearly established.

Although the vacuum facility was capable of reaching 10^{-9} torr, 3×10^{-7} torr was the lowest pressure (or highest vacuum) which could be reached with the specimen in the chamber within a draw-down time of 9 hr. This was more than double the time required to attain 10^{-6} torr. It may be significant that, as 3×10^{-7} torr was approached, outgassing of the material rapidly increased until it reached a rate equal to the draw-down rate of the equipment. Therefore, one is induced to associate this pressure level with the disruption of the last monomolecular layer of gas adsorbed onto the surface of the soil grains; at this stage, internal friction would be expected to increase substantially and possible local fusing or welding of the particles may take place. The occurrence of these phenomena would account for the significant increase in penetration resistance observed in Fig. 8 at vacuum levels of the order of 10^{-7} -10^{-6} torr.

To further evaluate this possibility two additional penetration tests were run at about 70 pcf using bake-out during draw-down and continuing the evacuation process for several days. In both cases vacuum in the 10^{-8} range was achieved, but as soon as the specimen was disturbed by the penetration the vacuum decreased substantially, reaching the 10^{-7} range. This indicates that the gases adsorbed in the specimen had not reached equilibrium with the chamber pressure. As a result there is no certain proof that the data at 70 pcf contradicts that at 65 and 75 pcf.

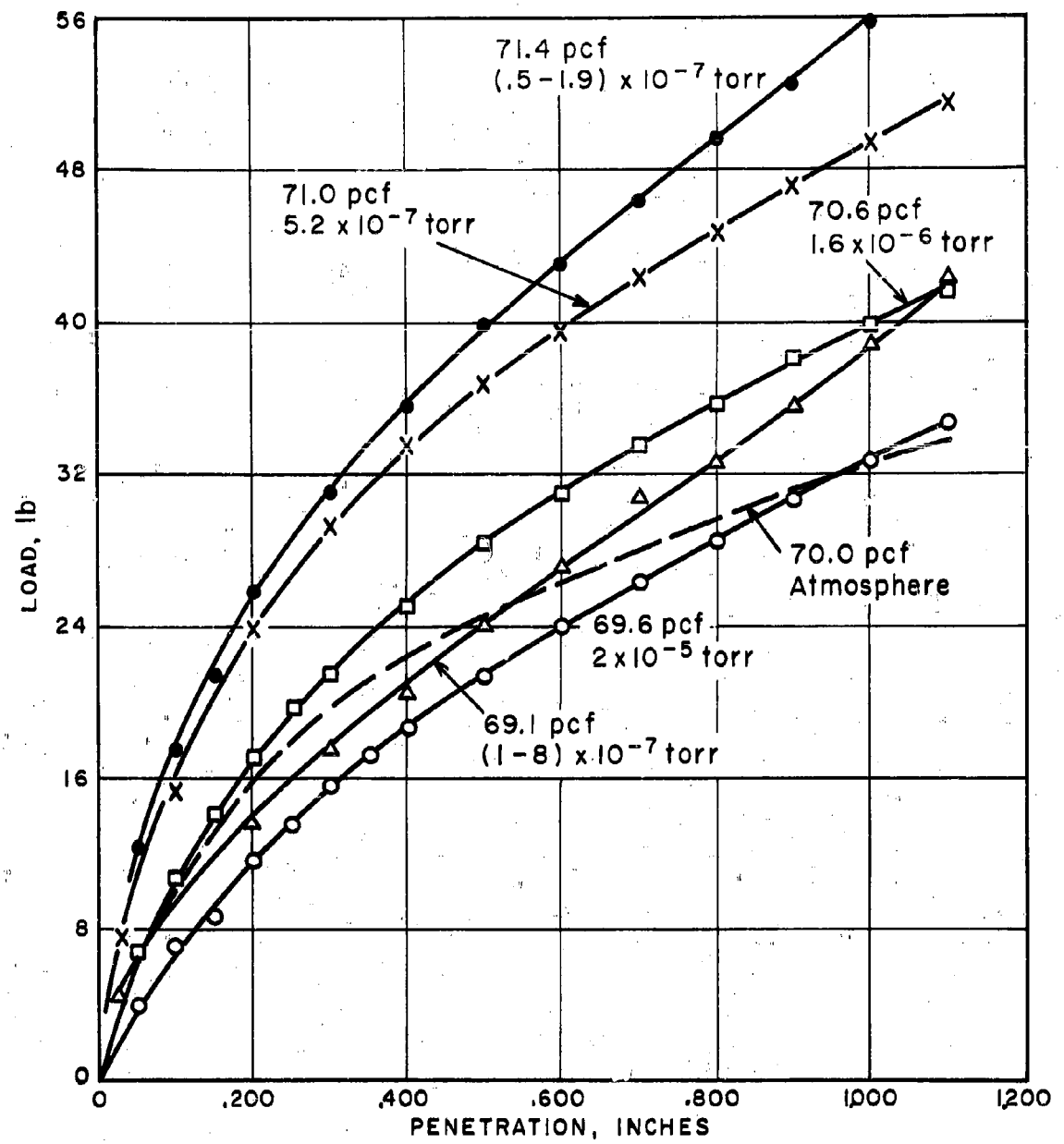


Fig. 7. Static Penetration Results in Vacuum Chamber for Specimen Density About 70 PCF

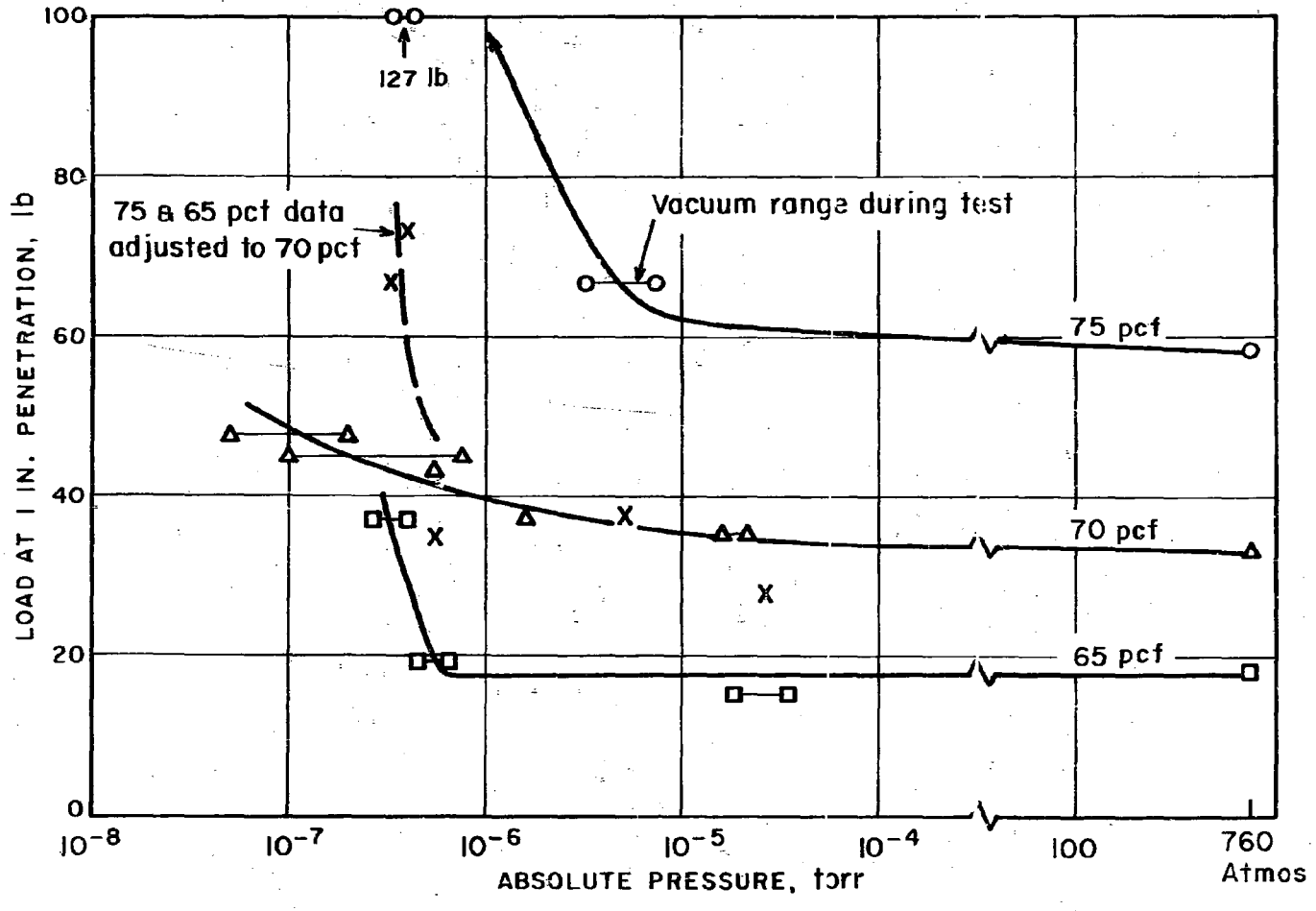


Fig. 8. Correlation of Static Penetration Resistance with Absolute Pressure

DYNAMIC PENETRATION RESULTS

The dynamic penetration test is the simplest one to perform which will give a relative measure of the strength of the material. The need to devise a loading system which will operate in the vacuum chamber is eliminated, and tests can be conducted at various vacuum levels without delaying the draw-down process. However, the phenomena involved in dynamic penetration are more complex than in static penetration. It was expected that both specimen density and vacuum level would influence the results, but, in addition, there are inertial effects and pore air pressures developed due to the rapid loading.

The major problem in correlating the results was that the penetrometer drop height and weight were not kept constant for the entire test series. The parameters defining the penetrometer include its weight W_p, cross-sectional area A_p, and height of drop H. To obtain a measure of penetration resistance and also provide a means of adjusting for small changes in W_p, A_p and H, a simple relationship was derived by equating the impact energy to the work done by the average resisting forces. In a rearranged form, this is

$$R_p = \frac{W_p}{A_p} \cdot \frac{H}{\delta} ,$$

where R_p is the average penetration resistance (force per unit area) and δ is the total penetration.

For this expression to be accepted as a valid measure of dynamic penetration resistance, it is necessary, among other requirements, that H/δ be constant for a given density. A series of tests were performed with all factors held constant except H to check this factor (Fig. 9). The relationship between H and δ was linear but showed a finite δ intercept. Hence the ratio H/δ changes with H, increasing as H becomes smaller. There was also considerable discrepancy found in the test results when either W_p or H was varied. Thus numerical correlation based on R_p was found to be valid only when all of the penetrometer parameters were held constant.

Effect of Absolute Pressure

The results of a series of dynamic penetration tests in the high vacuum chamber for several densities are given in Fig. 10. All of the points on each curve were obtained from at most two different specimens at essentially the same density. Except at the highest density (75.5 pcf) the data are all consistent and the scatter small. The dynamic penetrometer which has an area 1/14 that of the footing is more sensitive to local density variations, so that some scatter should be expected even when all of the readings are taken with the same specimen. The considerable variation of the data at 75.5 pcf. appears to be due to this cause, but, are accentuated because the total penetrations were very small and difficult to read accurately.

There is a definite trend for dynamic penetration resistance to increase with vacuum level. In the vicinity of 10^{-5} torr the increase is about 100% as compared to values corresponding to 1 atm. For vacuums above 4×10^{-7} torr

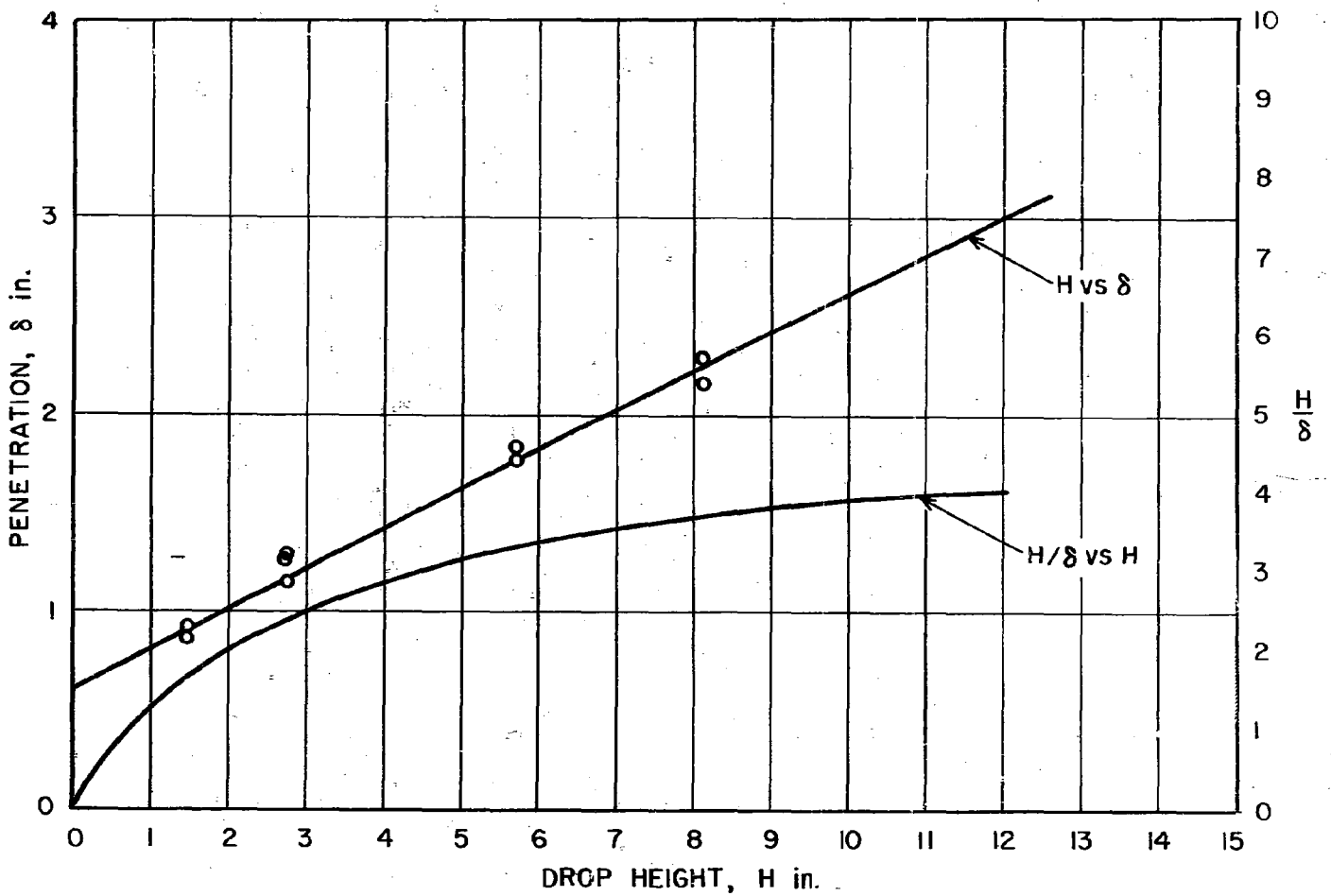


Fig. 9. Dynamic Penetration as a Function of Drop Height

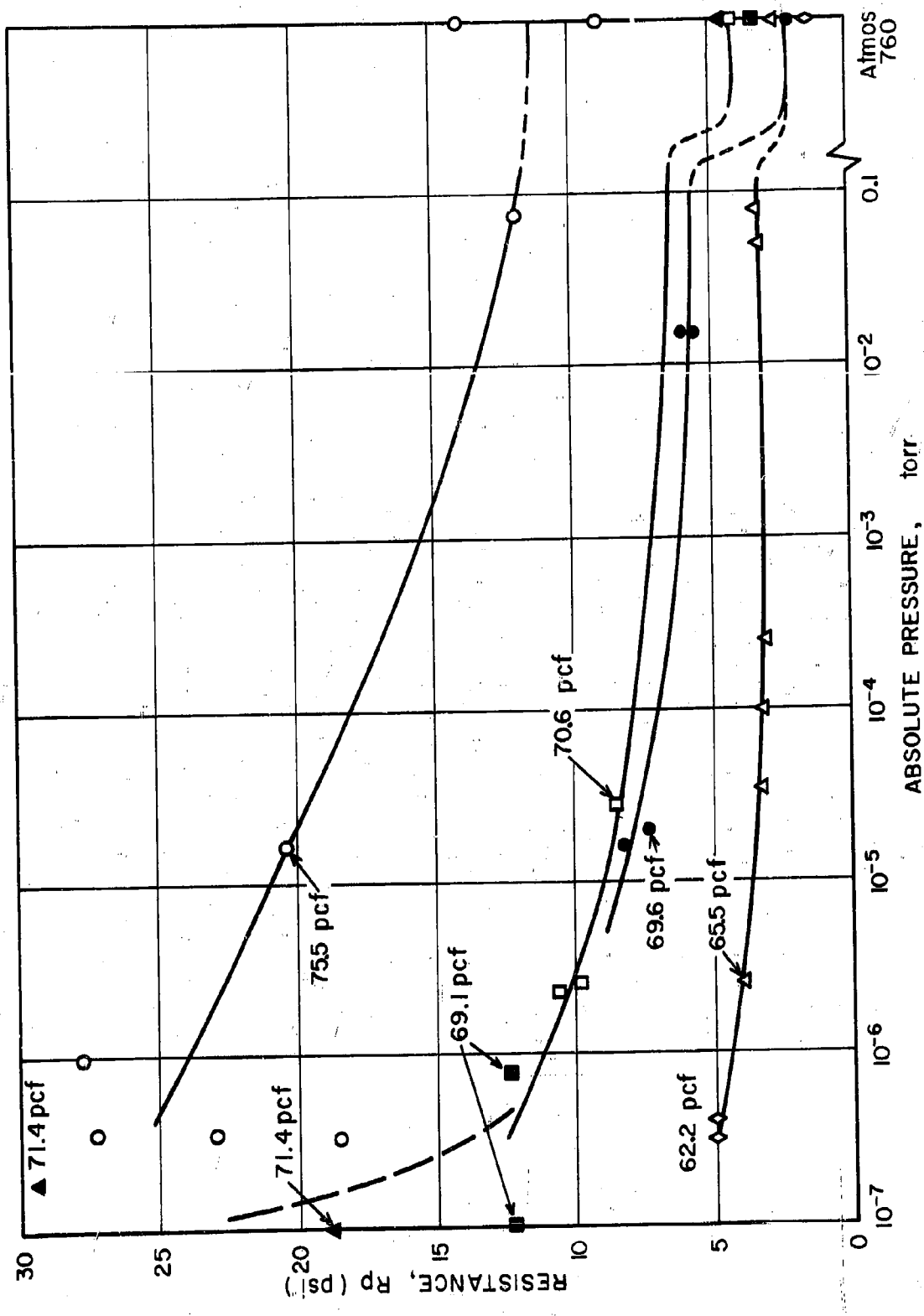


Fig. 10. Dynamic Penetration Results in Vacuum Chamber

an abrupt increase in penetration resistance is suggested by the results in the 70 pcf range. Probably by coincidence, this is the only density which did not show a significant increase in the static tests.

Pore Air Effects

In the pressure range of 12-760 torr, pore air pressures developed in the dust under the rapid loading create an additional influence on the penetration resistance by altering the effective stresses. The presence of pore air pressures can be observed from the manner in which the dust is displaced during penetration. Because the dust particles are very fine, there is considerable resistance to flow of the air from the voids as the dust is compressed. As the air tries to escape, a mound of dust is pushed up around the penetrating rod on the surface of the material (Fig. 4). In vacuum, when most of the air has been removed, no mound forms; instead, the dust is merely compressed beneath the rod. The mounding was evident in all of the bell jar tests down to a pressure of approximately 12 torr, but it progressively decreased with an increase in the amount of vacuum.

The influence of pore air pressures on penetration resistance was determined from tests in the bell jar (Fig. 11). As the vacuum level was increased the resistance gradually increased because, with the reduction of pore pressures, the interparticle forces increased. The penetration resistance at 12 torr was about 100% greater than that under atmospheric pressure.

SUMMARY

It has been clearly shown that both bulk density and absolute pressure are factors which significantly influence penetration resistance of finely ground silica powders. This material has been taken as one of these acceptably representative of the dust covering the surface of the moon if, in fact, such a dust cover actually exists. Other materials are expected to be drawn into investigation in the near future.

It is apparent that this material would have a strength under lunar environmental conditions which is higher than that which would be predicted based upon experiments under terrestrial atmospheric conditions. More data will be required in the high vacuum range to determine the exact degree of this increase in penetration resistance. The results suggest that at this pressure level the dust was in the process of a strength increase due to removal of the adsorbed monomolecular film of gas surrounding the particles. The dynamic penetration resistance, in addition, increases significantly and cratering characteristics change in the low vacuum range due to elimination of pore air pressures. Thus it is evident that representative dynamic tests cannot be performed on such fine materials under atomspheric conditions.

In this study penetration tests were chosen as a relatively simple method of evaluating the effect of density and vacuum on the strength of the silica powder. Such an evaluation must be largely qualitative since the strength parameters cannot be directly calculated from such penetration data. However, the penetration tests also can provide the basis for predicting the

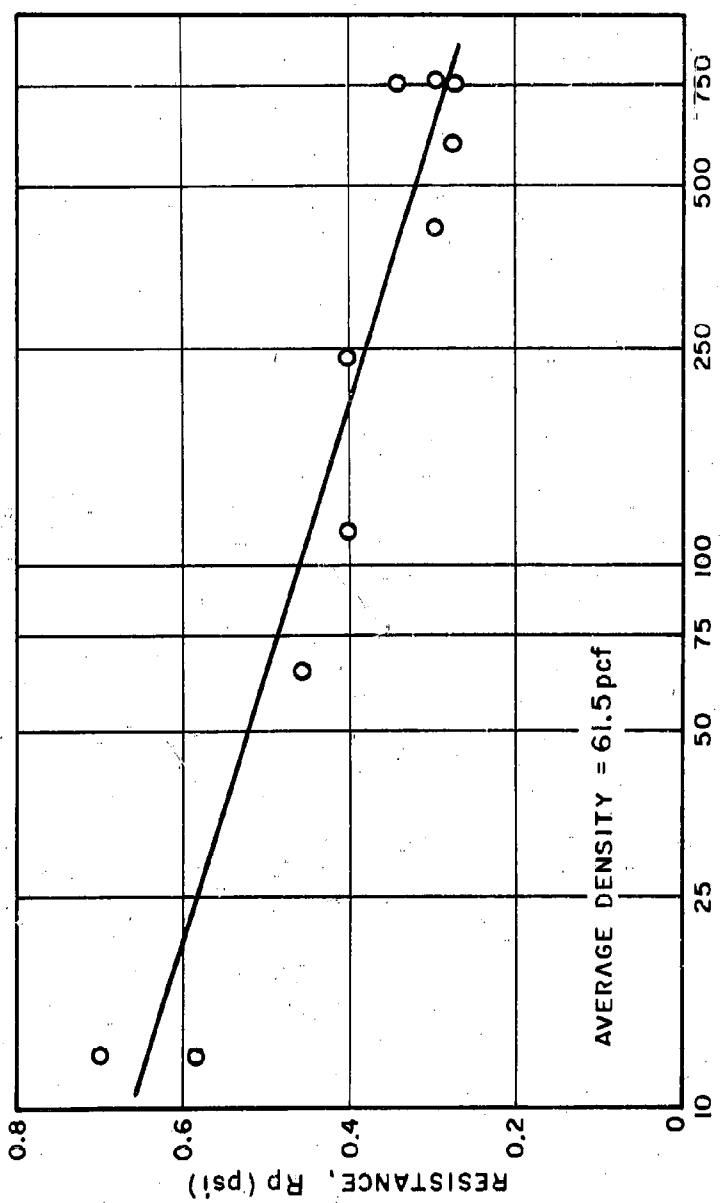


Fig. 11. Penetration Resistance in Low Vacuum Range

load-settlement behavior of various size footings as a function of density, vacuum level and layer depth. For instance, using a method of analysis investigated by Yong* it is estimated that the settlement for a corresponding load on a layer of infinite depth would be 30% greater than that for the finite layer used in the experiments.

Further studies are underway to fill some of the gaps in our knowledge of lunar soils. It is hoped that the experiments reported in this paper will serve to give a sound initial understanding of the behavior of these materials. It is believed that the results of such studies will make a significant contribution to the ultimate design of vehicles and equipment intended for operation on the lunar surface.

ACKNOWLEDGEMENTS

This study was sponsored by the Armour Research Foundation as part of its basic research program. The authors wish to acknowledge the interest and encouragement of the Foundation's Research and Invention Committee during the course of this research. The many helpful suggestions made by Dr. E. Vey, Dr. R. Yong and Mr. C. J. Costantino during the course of the work are appreciated. Mr. R. J. Larson is acknowledged for directing the operation of the high-vacuum facility. The authors would especially like to express their appreciation to Dr. N. A. Weil who originally suggested the program and whose inspiration and guidance have spurred on the efforts.

---

* Yong, R. N. Y., "A Study of Settlement Characteristics of Model Footings on Silt", Proceedings, First Pan American Conference on Soil Mechanics and Foundation Engineering, 1959.

STERILIZATION OF ELECTRONIC COMPONENTS OF SPACECRAFT

Joseph T. Cordaro
E. Staten Wynne

Aerospace Medical Division (AFSC)
USAF School of Aerospace Medicine
Astrobiology Branch
Brooks Air Force Base, Texas

ABSTRACT

Contamination of celestial bodies with earth microorganisms might make studies of any extraterrestrial life impossible. Sterilization of heat and/or radiation sensitive electronic components presents a special problem. Using a flexible film germ-free isolator, internal contamination was demonstrated in only 11 of 166 components, including 9 of 101 capacitors. Moreover, the level of natural contamination is low and destruction of microorganisms is assessed in terms of probability. Therefore, development of adequate sterilization procedures is being approached by deliberate contamination, during manufacture, with bacterial spores of high resistance to heat and irradiation. The results obtained with some types of resistors and diodes are presented.

STERILIZATION OF ELECTRONIC COMPONENTS OF SPACECRAFT

Joseph T. Cordaro and E. Staten Wynne
Aerospace Medical Division (AFSC)
USAF School of Aerospace Medicine
Astrobiology Branch
Brooks Air Force Base, Texas

INTRODUCTION

As pointed out by Lederberg (4), the best approach to direct studies of the possibility of extraterrestrial life involves microorganisms, since they are (1) an early stage in the evolutionary process, (2) ubiquitous on our planet and probably any other planet containing life, (3) easily cultivated, and (4) best adapted to automation and telemetry in unmanned space vehicles. However, such studies would be seriously hampered, or even made forever impossible, by inadvertent contamination with earth microorganisms deposited as a result of hard landings of space probes (1, 2, 4, 5, 7, 8) (figure 1). It is believed by some that earth microorganisms might well multiply on Mars (1, 2, 3). Such multiplication would not only interfere with determining the presence or absence of an indigenous flora, but by overgrowth might also jeopardize any possible use of extraterrestrial organisms for benefit of man (4).

Contamination of the moon with earth microorganisms would also jeopardize a most intriguing test of the panspermia hypothesis of Arrhenius, whereby spores drifted through space and seeded suitable planets. Having only a trace atmosphere, the moon has been envisioned as a trap for meteoroidal material (4), and is supposedly covered by a layer of cosmic dust (5). Although the surface temperature is estimated to range from +100 C. during the day to -180 C. at night, it has been estimated that at a depth of less than ½ meter, the temperature range is 0 to -70 C. Admittedly the ultraviolet radiation intensity on the surface of the moon is capable of killing all known microorganisms in a few hours, but Sagan (8) believes that microorganisms just beneath the surface, and thus protected from ultraviolet radiation, would survive cosmic radiation for a few hundred million years. Since the likelihood of a persistent indigenous flora is remote, the presence of microorganisms in the "moondust" would constitute evidence for the panspermia hypothesis.

Because of extremely low frictional resistance of the present trace atmosphere of the moon, it was recently stated by Phillips and Hoffman (6) that a single hard landing of a rocket could scatter bacteria over the entire surface of the satellite. As pointed out by Davies and Comuntzis (1), this possibility would be especially serious if a mammal were aboard, since there would be on the order of 10^{12} microorganisms present per kg of intestine. As the moon's surface is 4×10^{13} m^2, the microorganisms from one mammal might

"HARD LANDING" COULD DISSEMINATE EARTH MICROORGANISMS ON MOON AND JEOPARDIZE STUDIES OF:

1. EXISTENCE
2. NATURE
3. ORIGIN

} OF MICROBIAL LIFE

MICROORGANISMS MUST NOT BE PRESENT IN ANY PORTION OF SPACE VEHICLES, INCLUDING THE INTERIOR OF HERMETICALLY SEALED ELECTRONIC COMPONENTS. IRRADIATION OR STERILE FABRICATION MUST BE USED FOR HEAT-LABILE COMPONENTS.

Fig. 1. Hard Lunar Impact

give rise to a serious degree of contamination, especially since the density of any organisms trapped from cosmic infall would be expected to be small.

It has been pointed out (1, 7) that organisms might easily survive a space journey. Ultraviolet radiation in space, while admittedly lethal to any known microorganism in a few hours, would only be effective against organisms on the outside surface of a space probe. Furthermore, vacua actually aid in preservation of microorganisms, although it must be admitted that no data are available on the effects of vacua with residual pressures lower than 10^{-9} mm of mercury. Finally, it has been pointed out that entry into the atmosphere of Venus or Mars would not necessarily result in a probe being consumed by heat (1, 4). The atmosphere of Mars is mostly nitrogen, with only a trace of oxygen, while that of Venus is chiefly carbon dioxide.

In view of these considerations, there has been a great deal of concern over the question of contamination of celestial bodies with terrestrial microorganisms (1, 3, 4, 5, 7, 8). Conversely, the possibility of contamination of the earth with microorganisms from a celestial body is no less worthy of sober consideration. It is theoretically possible that microorganisms brought back from the moon or Mars might be capable of causing a new human, animal, or plant disease.

It would thus appear that rigorous measures are needed to insure the absence of viable microorganisms in a space vehicle prior to its launch. Davies and Comuntzis (1) have recommended sterile assembly, built-in disinfection and terminal sterilization. Sterile assembly may be aseptic or antiseptic. Aseptic assembly involves fabrication of sterilized components into subsystems, and sealing these subsystems to prevent recontamination. An example of antiseptic assembly would be the application of liquid wipe-on sterilants, such as formaldehyde in methanol, to all mating surfaces not accessible to gaseous disinfectant, e.g., nuts, bolts, screws, etc. Built-in sterilization might be useful in fabrication of certain components; e.g., paraformaldehyde might be included in plastic used for potting electronic components (7). In any case, terminal sterilization, particularly of the interior of any space vehicle, appears essential.

Suggested methods of terminal sterilization have included the use of (1) heat, (2) radiation, and (3) chemicals. A priori, heat sterilization appears most desirable. However, as emphasized by Davies and Comuntzis (1), the function of certain types of spacecraft components is impaired by commonly employed regimens using heat. With regard to terminal sterilization by radiation, formidable practical difficulties are involved, as pointed out by Phillips and Hoffman (7). Furthermore, Davies and Comuntzis (1) found that certain types of components would not tolerate sterilization by radiation. For terminal sterilization by chemical means, it appears that agents active in the vapor phase are most feasible. Of the gaseous disinfectants, ethylene oxide is generally considered the agent of choice for space vehicles (1, 4, 7).

Overall spacecraft sterilization procedures now employed include:

1. Internal sterilization of components.

(loose sand as before). Substituting these values in Eq. (9) gives R_s of 9.2 $lb_{\leftarrow}$ for 4 tracks per vehicle. To this we add R_h of 66.6 lb mechanical resistance Eq. (6) to arrive at R_{total} of 75.8 $lb_{\leftarrow}$, or 34 per cent of rigid wheel resistance.

Note that external resistance R_s of the track is only 12 per cent of the total, and that this value applies to an extremely weak soil. For longer tracks or slightly stronger soils, the external resistance can be neglected. It is thus evident that for tracked vehicles operating on frictional lunar surfaces, regardless of the soil strength, the rolling resistance is constant, determined for the most part by the mechanical, or internal, efficiency of the track.

Figure 2 compares the rolling resistances per pound for a rigid wheel, a flexible or pneumatic wheel, and a tracked vehicle, loaded with 250 $lb_{\oplus}$ as a function of wheel diameter or track length, while operating on the possible lunar surface extremes; a very weak frictional soil and a hard smooth surface. Figure 3 gives the same comparison for a 5000 $lb_{\oplus}$ load.

For the heavy load on assumed soft soil, the pneumatic wheel and track have equal resistances at 72-in. diameter and length. Slippage losses would be somewhat greater with the wheel; therefore, increasing the wheel to an 80-in. diameter should equate it to a 72-in. track. Increasing length of contact (D or S) beyond this point does not lower resistance significantly.

When operating on stronger soils, the flexible wheel is much more efficient than the track, particularly if wheel tread pressure is increased. As the weight of the vehicle is reduced, the advantage of the wheel becomes more apparent,

Figures 2 and 3 indicate that the rigid wheel cannot be made large enough without excessive size, weight and inertia to compete with a flexible wheel or track under the heavy load and assumed soft soil conditions; however, on a hard smooth surface the rigid wheel is the most efficient. With a lightweight vehicle operating on soft soil, the resistance advantage of a flexible tire or track is slight and a rigid or semi-rigid wheel might be favored for its simplicity and reliability.

Flotation on Soft Soil

If the vehicle does not deform the soil beyond its yield strength, it can be considered to float on the surface; in this case the resulting soil resistance is very low, approaching that of a hard surface. As can be seen in the previous analysis, the major part of the resistance of a rigid wheel would be eliminated, and the resistance of a flexible wheel would be reduced significantly but have negligible effect on the resistance of a track.

The maximum uniform static load W which a rigid surface may carry and maintain this flotation is given (5) by

$$W_{(max)} = 0.5 \gamma \quad b^2 s \, N_\gamma \quad \text{(dry frictional soil)} \quad (10)$$

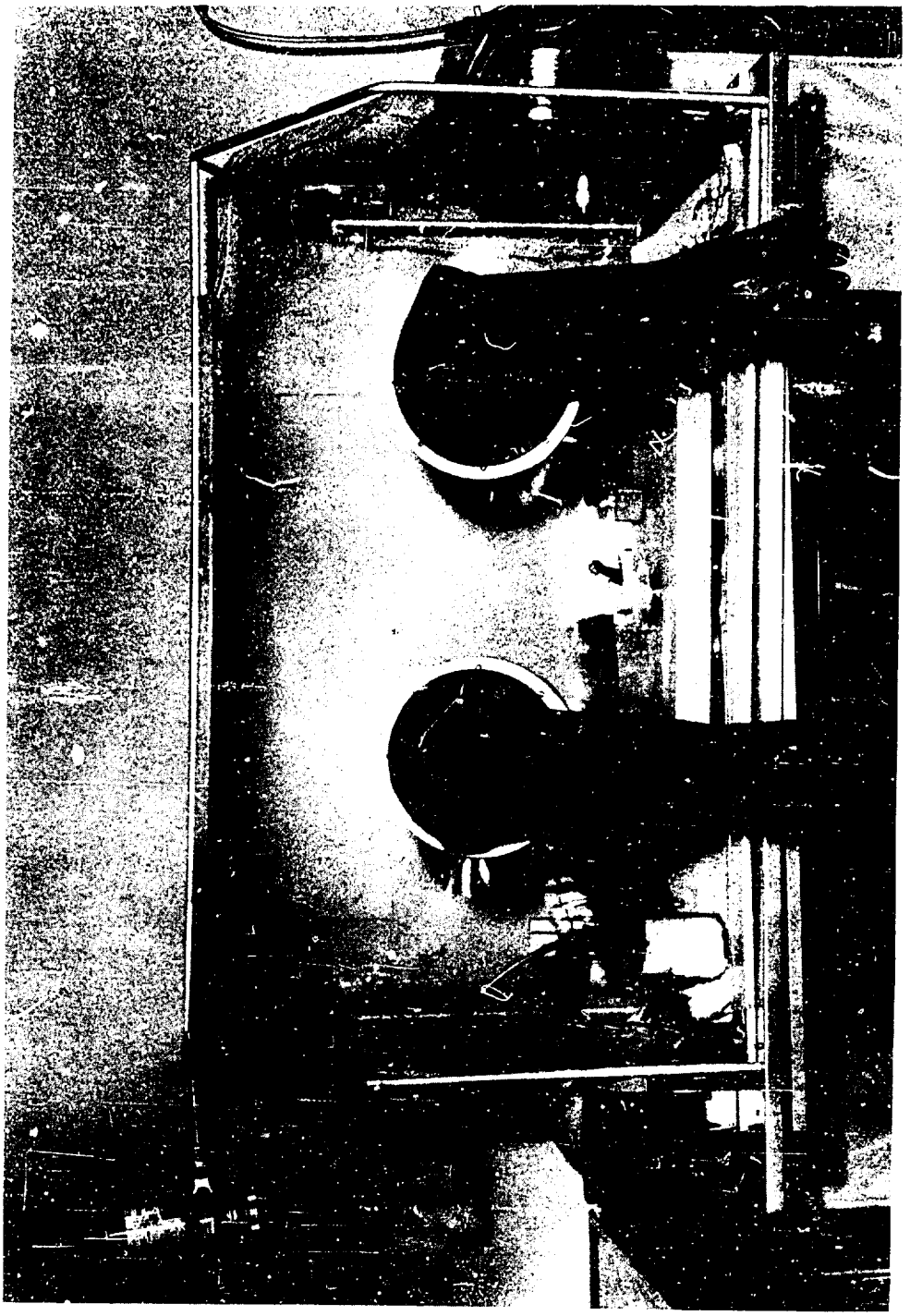


Fig. 2. Flexible Film Germ-Free Isolator

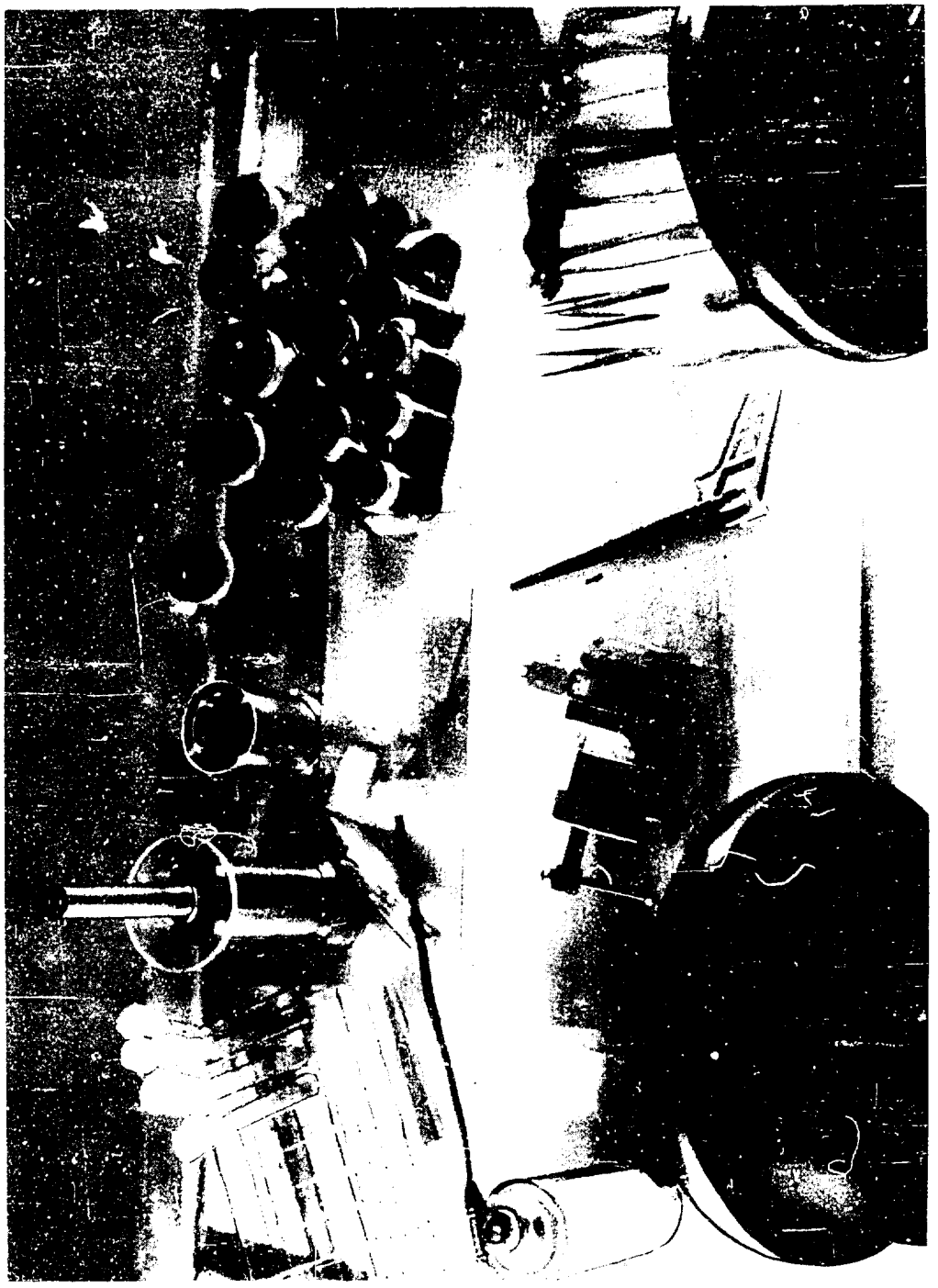


Fig. 3. Interior of Germ-Free Isolator

In view of these considerations, it is suggested that development of adequate sterilization procedures for sealed electronic components be accomplished by use of components which have been deliberately contaminated during manufacture. Such an experimental approach should include the following stages:

1. Selection of appropriate microorganisms with high resistance to dry heat and/or radiation.
2. Contamination, during manufacture, of the interior of components with standard high cell concentrations (1×10^5 to 1×10^6).
3. Performance tests of assembled contaminated components.
4. Exposure of contaminated components to dry heat time-temperature and/or radiation regimens.
5. Performance tests on heated and/or irradiated contaminated components.
6. Bacteriologic examination of undamaged components.

A joint in-house and contractual research program along the lines just detailed is in progress. Currently, spores of Bacillus stearothermophilus FS1518, suspended in acetone, are being used. The selection of FS1518 spores was determined not only on the basis of resistance to dry heat and radiation, but also to their ability to survive in a volatile liquid vehicle which does not damage the material to be contaminated. Recent data, however, by Koesterer and Bruch (9), and Davis et al. (10) seem to indicate that the spores of Bacillus subtilis var. niger are more resistant to heat than FS1518 spores. The resistance of B. niger spores to radiation and their survival in acetone is being investigated.

To date, four types of resistors (9 each) and two types of diodes (10 each) have been examined. The results indicate that for these types of components, normal manufacturing procedures are sufficient for sterilization.

SUMMARY

The importance of spacecraft sterilization has been discussed. Hermetically sealed electronic components constitute a special problem. Cultural techniques for such components have been described, and findings presented on the frequency of contamination. An experimental scheme for development of adequate sterilization procedures has been outlined, based on deliberate contamination during component manufacture. With the types of resistors and diodes tested to date, manufacturing procedures appear sufficient for sterilization.

Table 1. Results of the Different Types of Electronic Components Examined

Type of Component	Number Examined	Number Positive
Capacitors	101	9
Resistors	45	0
Diodes	5	0
Electronic tubes	5	0
Relays	2	0
Transformers	4	1
Magnetic modulator	1	1
Micropositioner	1	0
Potentiometers	2	0
Totals	166	11

REFERENCES

1. Davies, R.W., and M.G. Comuntzis, "The Sterilization of Space Vehicles to Prevent Extraterrestrial Biological Contamination." In Tenth International Astronautics Congress, Jet Propulsion Lab, External Publ. No. 698, Pasadena, Calif., 1959.

2. Davis, I., and J.D. Fulton, "The Reactions of Terrestrial Microorganisms to Simulated Martian Conditions," Proc. Tenth Int. Astronautical Cong., London, 1959.

3. Development of International Efforts to Avoid Contamination of Terrestrial Bodies, Science, vol. 128, 1958, pp. 887-889.

4. Lederberg, J., "Exobiology - Experimental Approaches to Life Beyond the Earth," Science, vol. 132, 1960, pp. 393-400.

5. Lederberg, J., and D.B. Cowie, "Moondust," Science, vol. 127, 1958, pp. 1473-1475.

6. Phillips, C.R., "Practical Aspects of Sterilization with Ethylene Oxide Vapor," Bacteriol. Proc., 23-24, 1950.

7. Phillips, C.R., and R.K. Hoffman, "Sterilization of Interplanetary Vehicles," Science, vol. 132, 1960, pp. 991-995.

8. Sagan, C., "Biological Contamination of the Moon," Proceedings of the National Academy of Science U. S. vol. 46, 1960, pp. 396-402.

9. Koesterer, M.G., and C.W. Bruch, "Resistance of Dry Bacterial Spores to Sterilization by Moist and Dry Heat," Bacteriological Proceedings Abstract A44, 1962.

10. Davis, N.S., G. J. Silverman, S. A. Goldblith, and W.H. Keller, "Survival of Spores at Several Temperatures in Ultrahigh Vacuum," Bacteriological Proceedings, Abstract A45, 1962.

PERFORMANCE AND ABORT CONSIDERATIONS AFFECTING THE SELECTION OF INJECTION CONDITIONS FOR EARTH - MOON TRAJECTORIES

Jim Rogers Thompson

Lockheed-California Company
Spacecraft Engineering
Burbank, California

ABSTRACT

The results of a brief study of the considerations affecting the selection of injection conditions for manned Earth-Moon Flights are presented herein. It is shown that injection conditions maximizing the system useful payload are considerably changed by the addition of abort provisions. The implications of the results on the selection of vehicle design characteristics and operational procedures are discussed through use of an example based on the Apollo system.

The work described in this paper was company sponsored.

PERFORMANCE AND ABORT CONSIDERATIONS AFFECTING THE SELECTION OF INJECTION CONDITIONS FOR EARTH - MOON TRAJECTORIES

Jim Rogers Thompson
Lockheed-California Company
Spacecraft Engineering
Burbank, California

INTRODUCTION

A rational design objective for manned space systems is the provision of capability to the pilot to abort the mission at any time and safely return to Earth. The stated NASA policy in this area is attainment of a probability of mission completion of 0.9, a probability of safe escape of .99, and a resulting probability of pilot survival of .999 for any conceivable failure of vehicle or operational systems. The provision of safe escape capability at suborbital and near orbital speeds has been the subject of extensive analysis (for example, references 1 and 2) and the design compromises associated with this capability are well understood and have been shown to be relatively small. At superorbital speeds, however, previous analyses (references 2 to 4) have shown that the penalty for provision of safe escape capability increases rapidly with increase in speed and at the critical point of injection to an escape trajectory, is critically dependent upon the injection conditions of altitude, velocity, and flight path angle. Further, the performance of the booster system is dependent on the injection conditions. In view of the above, the subject investigation was undertaken to provide a rational, quantitative basis for the selection of injection conditions as a function of vehicle design and abort system provisions.

The technical approach to the problem is first presented. Subsequent sections present the results of the investigation in generalized form (both from a performance and abort viewpoint) and discuss application of the derived information to a specific example taken from the Apollo system.

TECHNICAL APPROACH

The problem undertaken is that of obtaining quantitative evaluation of performance, abort provisions, and vehicle design and operational procedure considerations affecting the selection of the injection conditions for a manned Earth-Moon trajectory. Solution of this problem involves the following steps:

1. Select a typical mission description and representative Earth-Moon trajectories.
2. Determine the trajectories and injection conditions which maximize payloads for a typical booster vehicle and the payload penalities associated with nonoptimum operations.
3. Determine the requirements for abort to immediate reentry as a function of abort propulsion system design characteristics.
4. For a representative system, combine the payload and abort results obtained in steps 2 and 3 in order to determine the optimum design.

The following subsections discuss the basis and assumptions used in implementing items 1 to 3.

Selection of Mission and Representative Trajectories

The typical mission selected for the purposes of this study starts from Earth orbit at 100-nautical-mile altitude and is illustrated on Figure 1. The manned unit is carried by a single-stage, lox-hydrogen booster to injection into an Earth-Moon trajectory of the free return type which will reach the Moon in three days. The mission is considered to start from orbit rather than from the ground inasmuch as even single-shot systems such as Nova are generally considered to utilize a parking orbit for navigational reasons before proceeding to escape. The results presented, of course, are applicable either to a single shot or a rendezvous mission which proceeds from the specified orbit. It should be noted, however, that for certain launch vehicles direct ascent from Earth to injection might result in significant payload improvements. The Earth orbit altitude selected, 100 nautical miles, is believed to represent a practical compromise between the conflicting effects of the penalty for safe abort, which decreases with decreasing orbit altitude, and the provision of an orbit life sufficient for rendezvous techniques. Within the practical range of orbit altitudes, it is felt that change of the orbit altitude from 100 nautical miles would not result in significant changes in the general trends here presented.

The representative Earth-Moon trajectories used herein were selected from the results of reference 5, a parametric study of the trajectories resulting from a wide range of insertion conditions for lunar missions. Reference 5 shows that the trip time from injection near Earth to the point of closest lunar approach is primarily a function of the incremental velocity

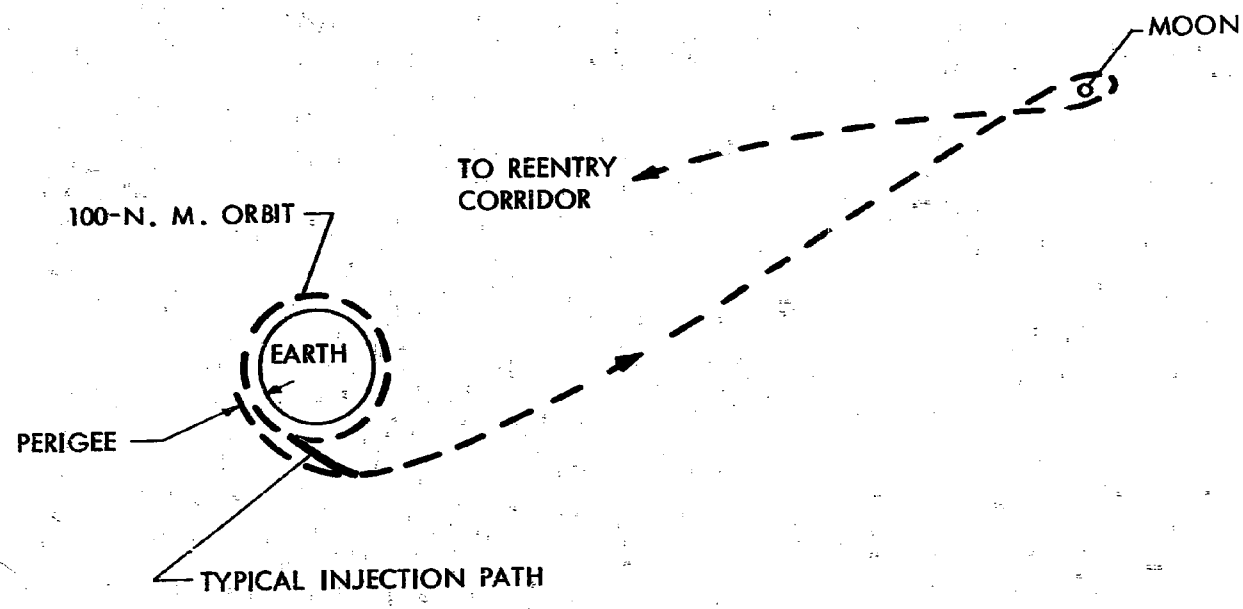


Fig. 1. Selected Mission

at injection above that required to just reach the Moon's sphere of influence. An incremental velocity of about 100 fps provides the desired three day trip. It is also shown that for incremental velocities greater than about 30 fps, an injection lead angle exists which provides both a small lunar miss distance and a free return to within 40 miles of the Earth. It is thus apparent that the incremental velocity of 100 fps meets the trip time and free return requirements of the specified mission. The lunar lead angle is not considered further in this study inasmuch as it has no effect on performance or abort, but only specifies the time and location of injection.

Because injection at either high altitudes or high flight path angles can be immediately shown to result in large penalties for abort, principal emphasis in this study was placed on trajectories injecting at low altitudes and low angles. Within the constraints discussed, five representative Earth-Moon trajectories having incremental velocities of 100 fps were selected from the data of reference 5. The perigee altitudes ranged from 316,800 to 1,000,000 feet. The results presented subsequently will show that this range covers the region of practical interest. The variation with inertial velocity of the altitude and the flight path angle measured with respect to local horizontal of two of these Earth-Moon trajectories are shown on Figure 2.

Abort

There are two basically different procedures which can be utilized to provide safe abort at any point of the trajectories shown in Figure 2. One of these involves firing a sufficient retro impulse to reduce the apogee of the trajectory from the vicinity of the Moon to near the Earth. Previous analyses have usually discarded this method on the basis that reduction of the apogee sufficiently to provide entry in a reasonably short period of time results in trajectories which pass through critical regions of the Van Allen radiation belt and thus imply high dose rates. The other method is that of applying the abort propulsive impulse approximately toward the center of the Earth so as to reduce the flight path angle to a value such that the atmosphere will be intercepted at an entry angle below the skip out limit, thus allowing immediate reentry. A representative skip out limit near escape speed for systems such as Apollo operated at maximum downward lift is about 6 degrees below the local horizontal measured at an altitude of 400,000 feet. The variation with velocity of the altitude and flight path angle for an entry trajectory resulting in this limiting entry condition is shown on Figure 2 by the dashed lines.

The velocity vector diagram associated with abort from an Earth-Moon trajectory is also presented on Figure 2. It is immediately apparent from consideration of the vector diagram and of the trajectory data presented on Figure 2 that injection at higher flight path angles (higher altitudes and lower velocities) results in rapid increase in the amount of angle change necessary for immediate reentry and, therefore, rapid increase in the propulsion required for abort. It is also apparent that because the abort propulsion vector is supplied approximately normal to the flight path direction, small differences between the velocity along the reentry path and

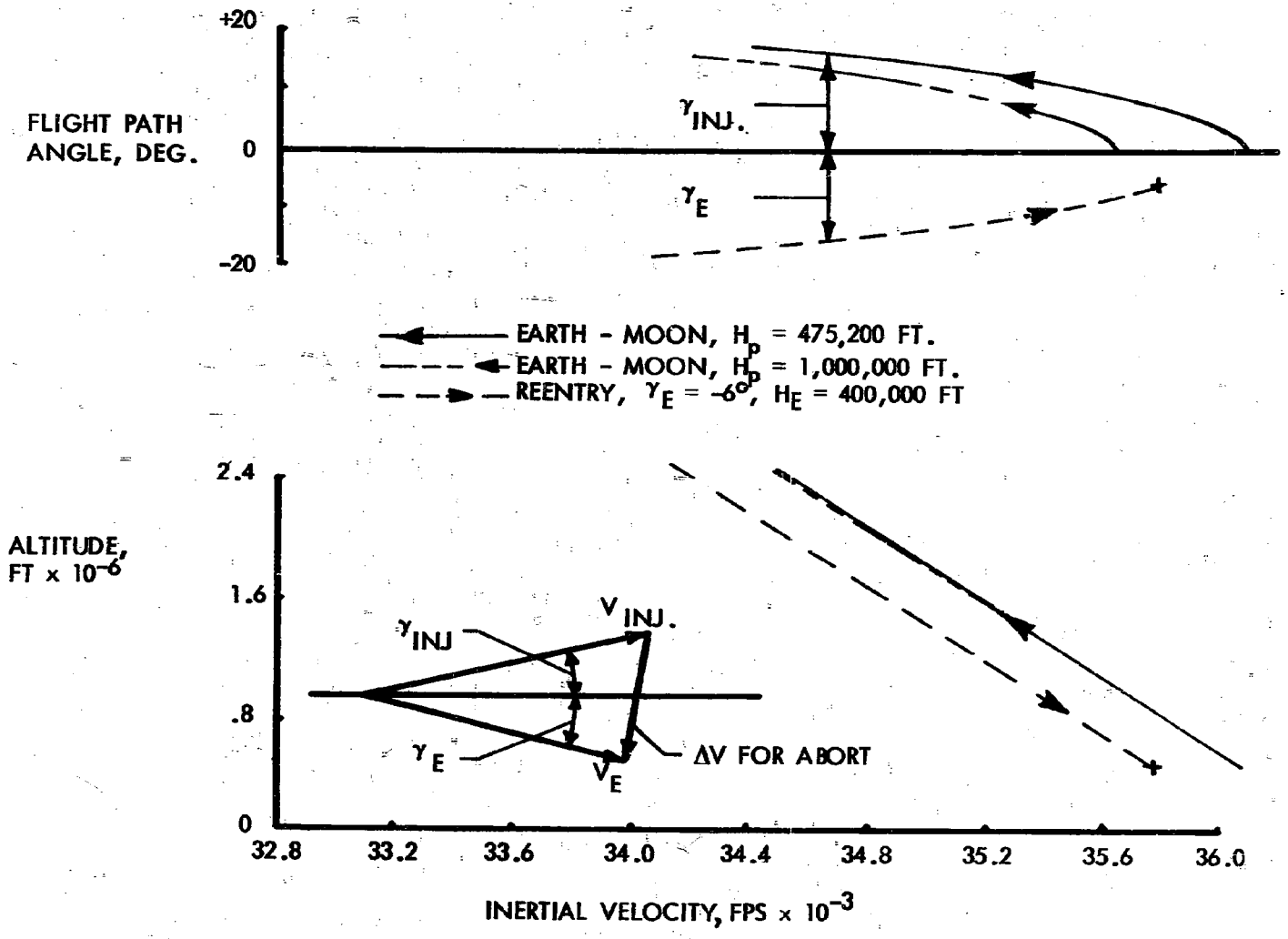


Fig. 2. Earth-Moon and Reentry Paths

the Earth-Moon trajectory can be corrected at negligible propulsion cost by tilting the abort velocity vector away from the normal. For this reason the entry velocity value used is representative of a relatively wide range of entry velocities although the entry angle is critical.

Performance Optimization

The optimum performance of the launch vehicle from orbit to injection conditions along the Earth-Moon trajectory was determined through use of the Lockheed HI-STEP (High-Speed Integrated Space Transportation Evaluation Programs) which, by means of an iterative procedure performed on a digital computer, determines the optimum trajectory and angle of attack program which both satisfies the desired end conditions and maximizes the payload of a specified launch vehicle. A variant of this program which has an additional variable of optimization, a coast time after burn out, was used to determine the optimum injection conditions for the specified vehicle. This result was accomplished by utilizing an end condition far beyond the expected injection point and thus optimization of the coast time from injection to the end point provided the optimum point along the trajectory at which the injection should be accomplished.

RESULTS

The study results are presented in the following subsections. The first two subsections present the performance and abort propulsion requirements, respectively. The concluding subsection discusses the application of these data to an example system.

Performance

The performance of the selected single stage, lox-hydrogen fueled vehicle between 100-n.mile circular orbit and each of the several Earth-Moon paths was first determined without regard to the abort problem. The resulting performance map is presented on Figure 3 as the variation with perigee height of the Earth-Moon path of the initial-to-burn-out-weight ratio for two injection conditions and three initial thrust-to-weight ratios.

The limit line for infinite T/W was computed directly from the difference between the orbital velocity vector and the velocity vector of the Earth-Moon trajectory at the orbit altitude. The payload decreases with decrease in perigee because of the increase in flight path angle at the orbit altitude as the perigee is reduced below the orbit altitude. The data for finite T/W approaches the infinite T/W limit as the perigee is reduced.

The lines labeled "optimum" were obtained by free-coast runs in which the HI-STEP program selected both the optimum angle-of-attack program and the optimum point of injection to each of the Earth-Moon trajectories. The injection points determined in this manner for T/W of .47 ranged from 1.2 to 1.6 x 10^6 feet at flight path angles about 10 degrees; and for T/W of .94, from .6 to 1.1 x 10^6 feet at angles near 5 degrees.
The results show that the maximum payload (lowest weight-ratio) is obtained by

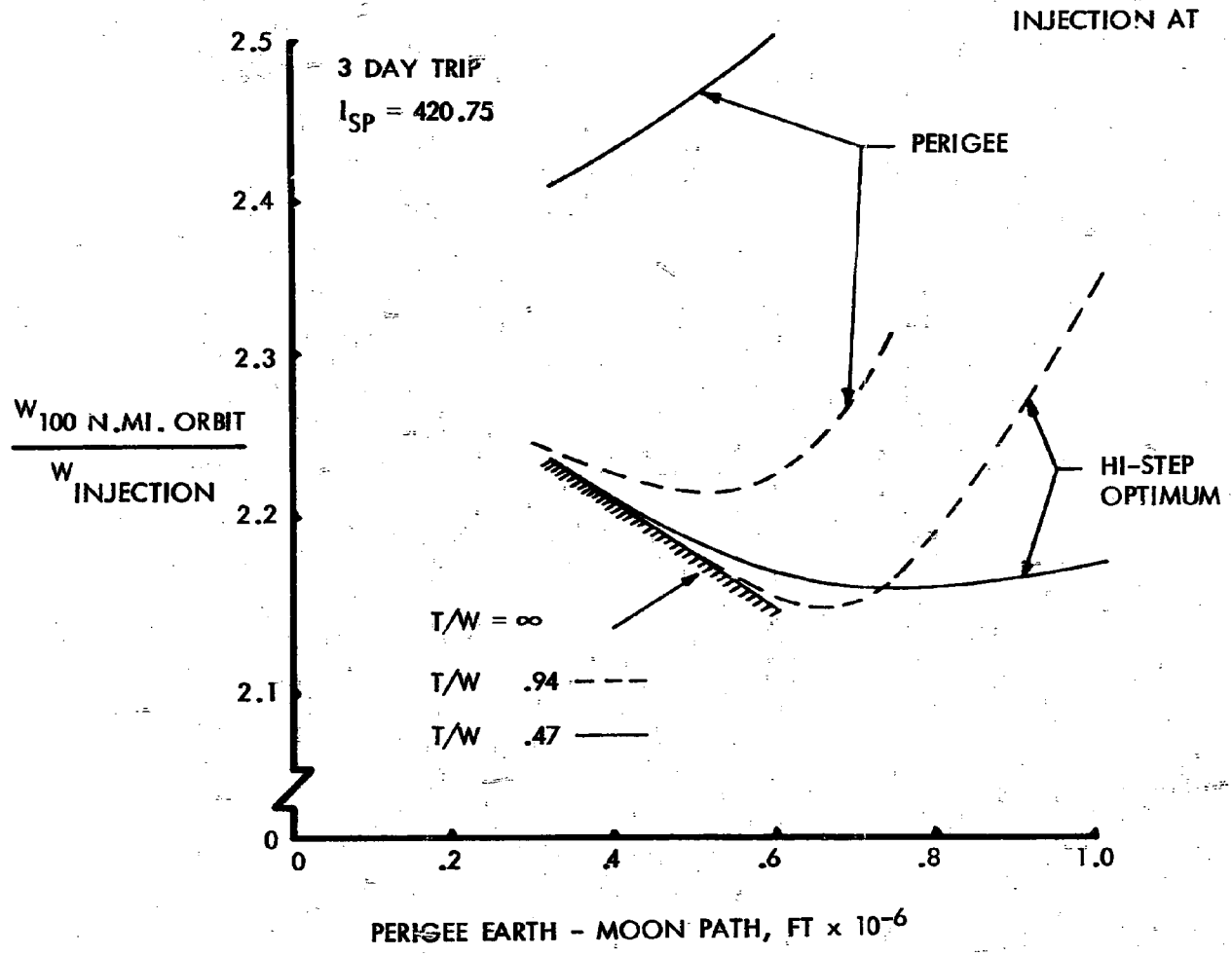


Fig. 3. Injection Performance

injection to an Earth-Moon path having a perigee greater than the orbit height by an amount which increases as the T/W is reduced.

The lines labeled "perigee" were obtained from HI-STEP optimum angle-of-attack program runs to the specified perigee points without using the free-coast feature. It is immediately apparent that injection at perigee results in a considerable reduction in payload, and that the maximum payload occurs for lower perigees than in the case of optimum injection.

The effect of injection at conditions between optimum and perigee is illustrated by Figure 4 which shows a cross section of Figure 3 at a perigee of 475,200 ft. The curves are flat over a wide range of injection angles and significantly reduced payloads occurs only for injection near perigee at the lowest value of T/W shown. Reference to Figure 3 shows that the penalty for perigee injection increases with increase in perigee above that of Figure 4, however.

Figures 3 and 4 show that if the abort problem is ignored, an Earth-Moon perigee above the orbit altitude should be selected and injection performed at relatively high altitude in order to maximize payload. In order to explore the effects of abort provisions on this result, two types of missions differing in propulsion requirement after injection must be considered. In the first type, such as a lunar fly-by, the normal mission does not require large amounts of fuel after injection and thus provision of the propulsive impulse for abort at injection is a direct penalty. The penalty for abort from injection to immediate reentry has been computed for infinite thrust and a specific impulse of 315 through use of the abort characteristic velocity determined in accordance with the vector diagram insert in Figure 2. The variation with injection angle of the overall orbit-to-reentry-weight ratio (the product of the abort penalty weight-ratio and the orbit-to-injection ratio taken from Figure 4) is presented on Figure 5 labeled "Minimum Abort". The figure shows that within the limitations of the infinite-abort-thrust assumption, injection for this type of manned mission should be at or before perigee in order to maximize useful payload.

The second type of mission is one in which considerable propulsion requirements exist after injection; for example, lunar landing. The trade-off curve representative of this type of mission is obtained by multiplying the orbit-to-injection-weight ratio data of Figure 4 by the injection-to-reentry weight ratio for the normal mission fuel. Data are shown on Figure 5 for an injection-to-reentry weight ratio of 2.765. This ratio corresponds to a characteristic velocity of 10,500 fps, approximately that of the Apollo service module. The figure shows that for this condition, net payload is maximized by injecting at the maximum angle for which the available propulsion provides abort capability. For the example shown, this point is about 5 degrees for infinite thrust and would be appreciably lower for finite thrust.

In order to provide a summary of the trends indicated by this study, the overall weight-ratios from Figure 5 and similiar data for other perigees are plotted against perigee height on Figure 6 for injection at perigee and at 3 degrees. The general trend of the data is toward higher payload at the lower perigees. As perigees below 400,000 feet might encounter

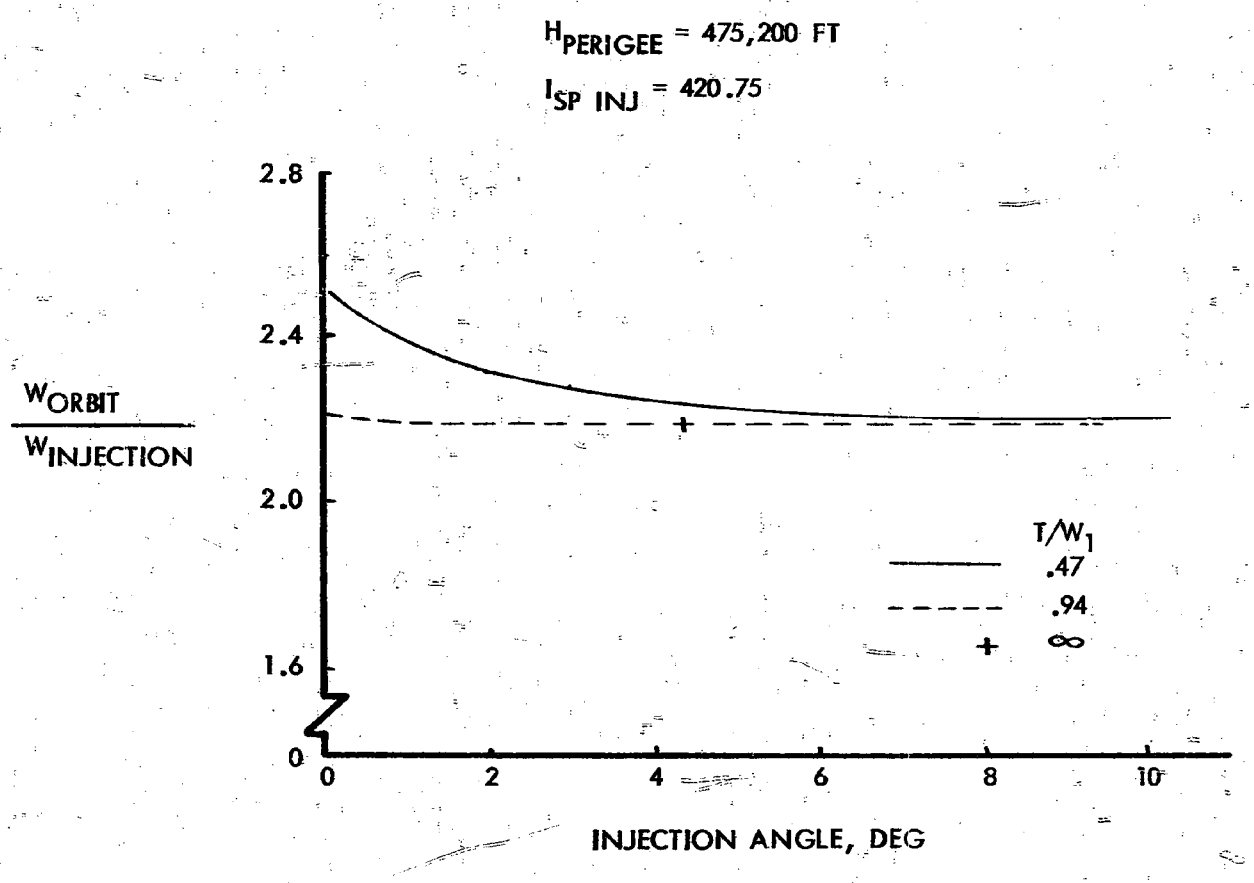


Fig. 4. Effect of Injection Angle on Performance

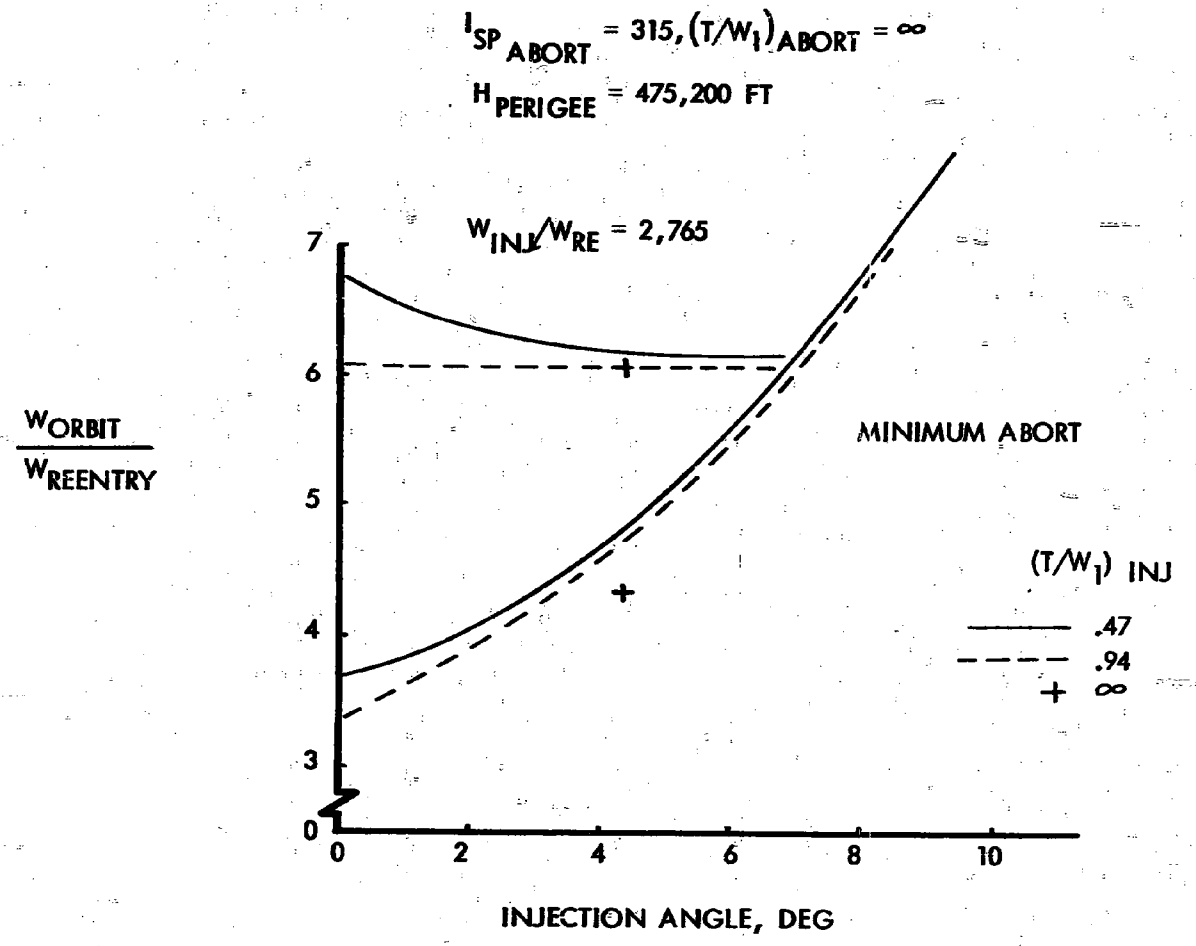


Fig. 5. Effect of Abort on Performance

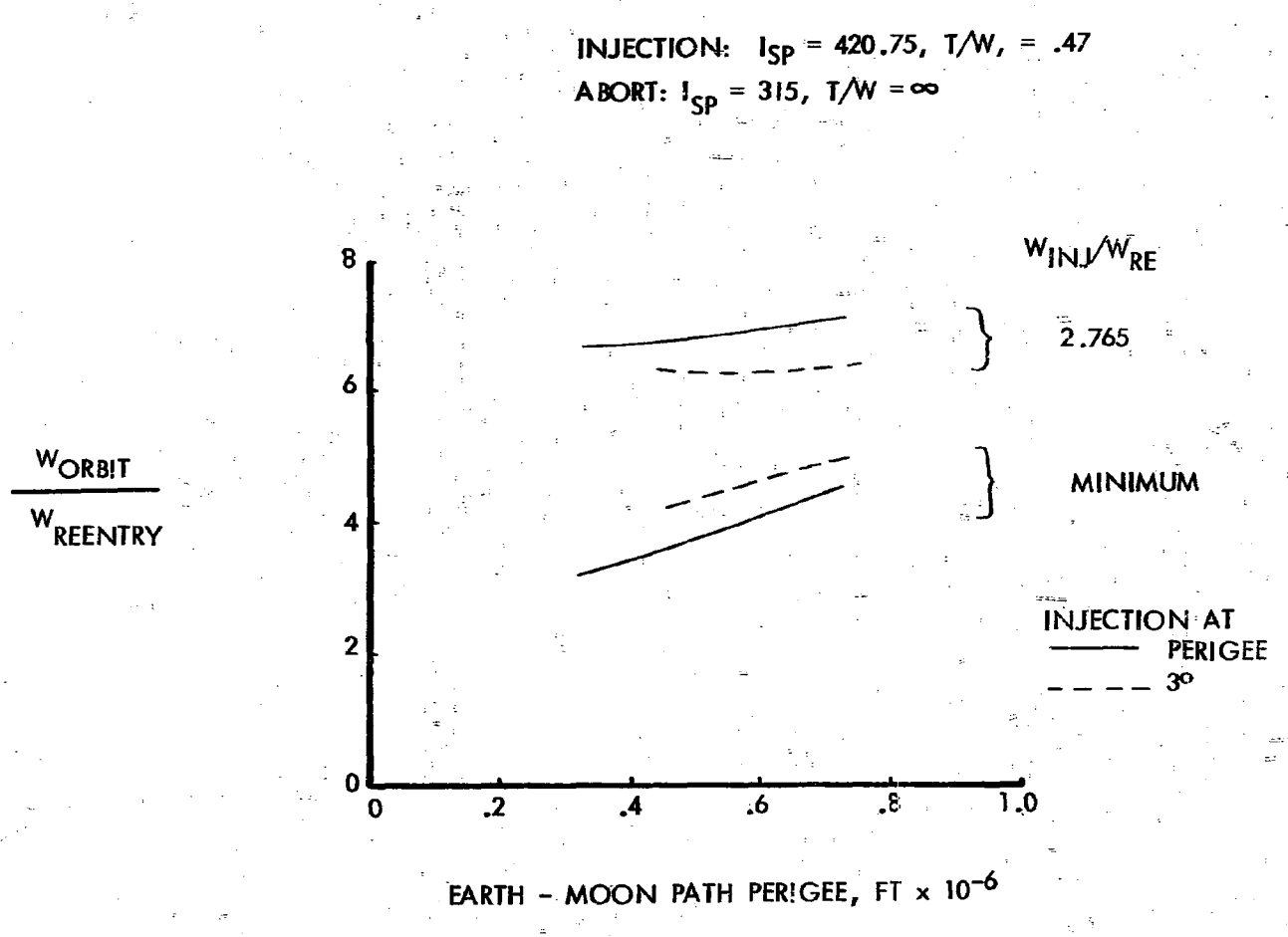


Fig. 6. Summary of Abort Effects on Performance

significant aerodynamic heating, a perigee of 475,200 feet appears to be a reasonable compromise and was selected for use as an example herein.

Abort Propulsion Requirements

It was discussed in connection with Figure 2 how the impulse requirement for abort from any point on the Earth-Moon trajectory to the associated point on the reentry trajectory (also shown on Figure 2) could be determined through use of the velocity vector diagram. An additional factor is involved, however, inasmuch as near escape speed the centrifugal force on the vehicle is approximately equal to twice the vehicle weight and the system is effectively operating in a negative 1-g field; thus, an abort thrust-to-weight ratio of 1 would require infinite impulse, and a "gravity loss" is encountered which is proportional to the thrusting time of the abort system. For this system the initial-to-burn-out weight ratio of the abort propulsion system is approximately (1)

$$W_1/W_2 = \ln^{-1} \left\{ (\Delta V_o + g t_b)/I_{sp} g \right\} \qquad (1)$$

where

ΔV_o = velocity increment from vector diagram of Figure 2

t_b = abort propulsion burning time

I_{sp} = specific impulse

g = acceleration of gravity

It can be easily shown that the abort system thrust-to-initial weight ratio of the vehicle (for a constant thrust) is given by (2)

$$T/W_1 = \frac{I_{sp}}{t_b} \quad (1 - W_2/W_1) \qquad (2)$$

Utilizing these relationships the variation of characterisitc velocity of the propulsion system required for abort to immediate reentry from injection with thrust-to-weight ratio has been computed for the injection conditions considered in the performance analyses. Results of these computations are presented on Figure 7. The characteristic velocity is plotted against the inverse of the thrust-to-initial-weight ratio so that the infinite thrust (zero burning time impulsive case) can be shown. Lines of constant burning time are also shown on the figure. Examination of Figure 7 reveals that for the impulsive thrust case, from 4,000 to 14,000 ft/sec of characteristic velocity is required of the abort propulsion system and that for a fixed injection condition the characteristic velocity required for abort increases rapidly with decrease in thrust-to-initial-weight ratio.

As an example of the application of this figure, a point is shown near the center of the figure labeled Solid Service Module, T_b = 50 sec. This point refers to the service module of the Lockheed-Chance-Vought-McDonell proposal operated in such a manner that all five solid rockets are burned

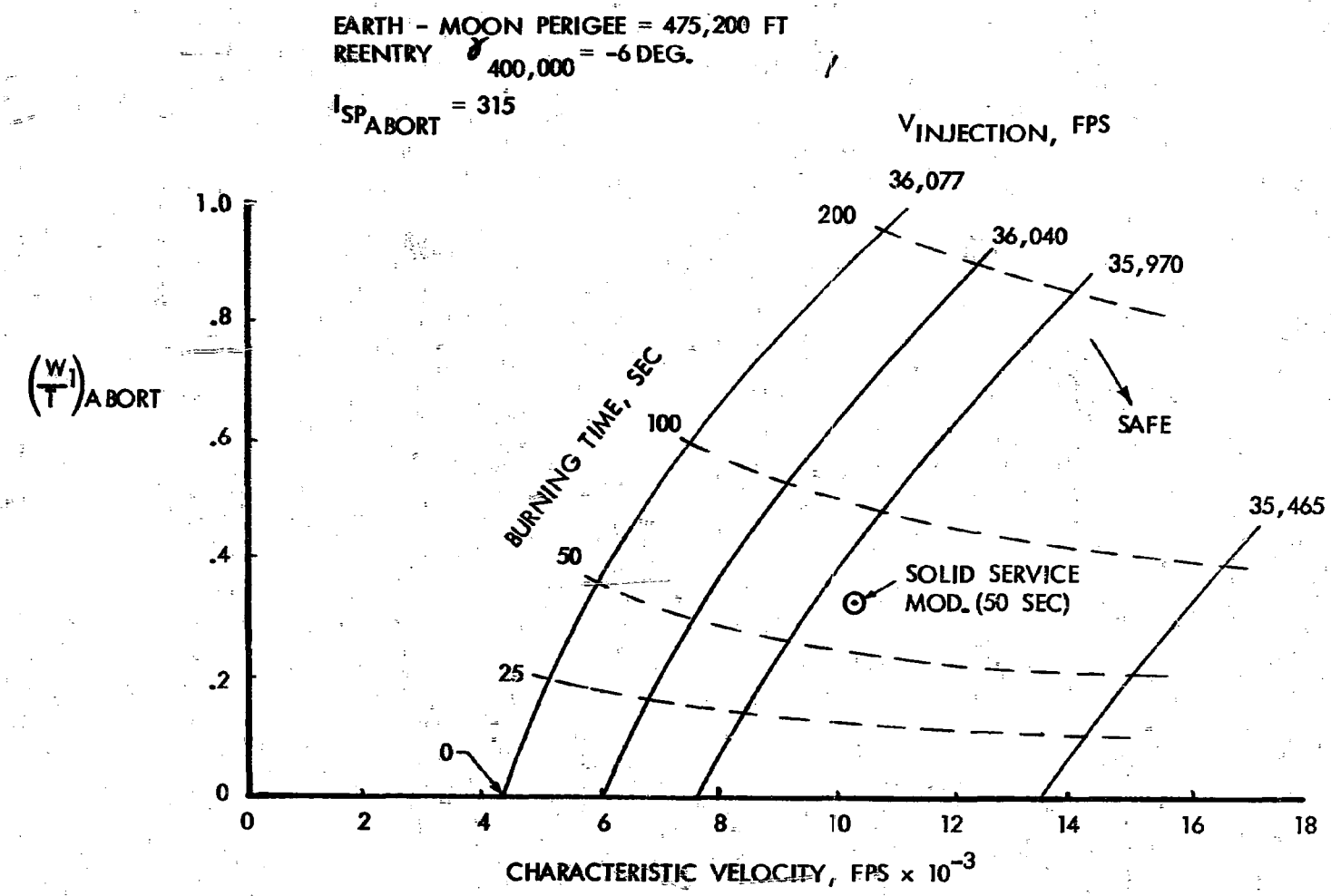


Fig. 7. Propulsion for Abort at Injection

simultaneously together with the vernier system thus giving a thrust-to-initial weight ratio of about 3. This system has a characteristic velocity potential of about 10,300 ft/sec which exceeds that required for safe abort at injection speeds above about 35,950 ft/sec. However, it is apparent that the point lies above the line for 50 sec burning time; thus it has insufficient thrust-to-weight ratio to perform an abort from this condition. If the point were moved to the left at constant thrust-to-weight until it is under the 50-sec line it is apparent it would be capable of safe abort at injection speeds of about 36,060 ft/sec; however, it would require only about 6,500 ft/sec characteristic velocity for this abort and thus has about 30% excess fuel.

The performance of the abort system operated in the retro mode discussed earlier was briefly explored in order to provide a comparison with the abort-to-immediate-reentry mode. It was found that from perigee injection conditions, firing a 10,300 ft/sec retro impulse with a thrust-to-initial weight ratio of one resulted in an apogee of 425 n.miles and reentry at about -6 degrees near orbital speed. Reentry started about 3,000 seconds after abort; about ten times the delay associated with abort to immediate reentry. Comparison of the retro case with Figure 5 shows that the impulse requirements of the two methods are similar at the conditions investigated.

It was also found that in the retro mode, reduction of abort system velocity increment, thrust-to-weight ratio, or injection velocity (increase in injection angle) resulted in rapid increase in apogee above 500 n. miles. Further, reduction of T/W or injection velocity resulted in rapid increase in entry angle. Obviously, the acceptability of high apogees and steep reentry angles depends on system design characteristics such as shielding, lift capability, thermal protection, etc. From the results of the limited exploration presented it appears that the retro abort mode requires impulse and T/W's similar to those required for abort to immediate reentry and involves, in addition, a delay to reentry an order of magnitude greater. Thus, the retro mode does not appear promising. A considerably more extensive study, including combined modes, is of course required to document the above conjecture.

Discussion and Application

In order to show the effect of the performance and abort requirements on a practical system design, an Apollo mission utilizing the S-IVB Saturn stage has been selected as an example. According to current public information, the Douglas S-IVB stage has a burnout weight of 26,000 lbs and carries 198,000 lbs of propellant. The stage is equipped with a single J-2 engine giving an initial thrust-to-weight ratio of about one-half; thus the performance data presented previously for T/W of .47 is closely applicable. For purposes of this example the S-IVB stage is assumed to be fully fueled in an 100-mile Earth orbit. Its payload consists of an Apollo command module, a service module having an I_{sp} of 315, and a lunar landing module. The distribution of weight between the various components has been computed assuming that the service module fuel capacity varies so as to supply the abort propulsion requirement. The remaining payload is considered to be that available for the lunar landing module. The variations of the weight

distribution computed in this manner with injection velocity are presented on Figure 8. Inasmuch as the S-IVB fuel and inert weights are independent of the injection condition the variation of S-IVB payload with injection velocity directly reflects the performance curve already presented on Figure 4. The payload thus delivered decreases from about 145,000 lbs at the optimum injection condition (35,465 ft/sec, 10.3 degrees) to about 115,000 lbs for injection at perigee. The service module fuel required for abort to immediate reentry (assuming an abort-thrust to initial-weight ratio of 2-1/2) increases rapidly with reduction of injection velocity. Although not evident at the scale of the figure, the service module inert weight is also changed to account for a tank weight of 5 percent of the service module fuel weight. After addition of the constant command module weight, it is apparent that the payload weight available for the lunar landing module varies from about 30,000 lbs at the "optimum" injection condition for S-IVB from a payload viewpoint only, to a maximum of about 90,000 lbs at an injection velocity of about 36,030 ft/sec. This optimum condition corresponds to an injection angle of about 2-1/2 degrees above the local horizontal. For further reduction in injection angle the performance reduction overpowers the reduction in abort penalty with the result that the net payload available for lunar landing is reduced.

In the Apollo mission concept the service module is required to provide the propulsion for takeoff from the lunar surface and injection into the Moon-Earth trajectory as well as provide the abort requirements. Examination of the mission requirement shows that the service module must for this purpose have a characteristic velocity potential of about 10,500 ft/sec. The service module shown on Figure 8, designed from abort requirements alone, has shrunk to that size at an injection velocity of about 35,960 as indicated on the figure. If the service module is kept constant at that size then the amount of payload available for the lunar landing module is that shown by the dotted line. For this approach the optimum payload available for the lunar landing module is about 88,000 lbs and occurs at an injection velocity of about 35,960 ft/sec.

It should be noted that the thrust level desirable for the service module is drastically different for the abort application and the mission application. As shown in Figure 7, increase in T/W for the abort segment results in rapid reduction in the service module size, for example, reduction of abort T/W in the case shown from 2-1/2 to 2 would result in an increase fuel requirement of about 11 per cent. For the lunar take off case however, T/W should be appreciably less than 1 in order that a large payload penalty not be encountered. The payload penalty for lunar takeoff at high thrust can, of course, be avoided by use of a coast period. However, use of a coast in the lunar takeoff implies an engine restart with attendant reduction of reliability. From the above it appears that the service module design should probably include at least two engines, a high thrust engine for the provision of abort capability at injection and lower thrust engines which could be operated alone for the lunar take off and guidance impulse requirements of the service module. It might be possible to design the system, for example, similar to the Atlas launch vehicle in which after injection to the Earth-Moon trajectory the high thrust engines could be jettisoned so that they would not penalize the

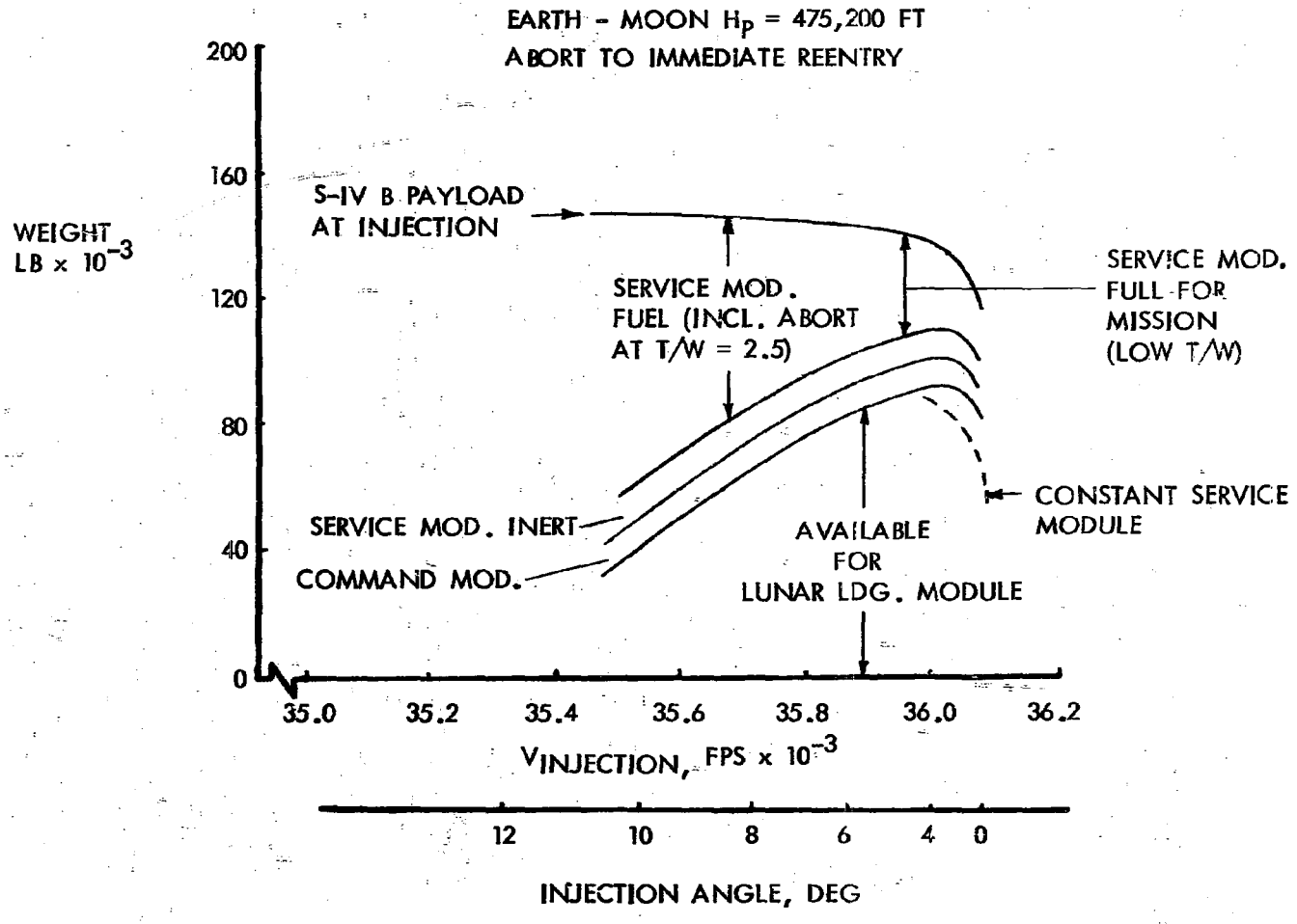


Fig. 8. Example of Injection and Abort Effects

performance of the lunar landing and service modules in the mission segments near the Moon.

With regard to abort at conditions other than the instant of injection into the Earth-Moon trajectory, two cases are possible. One, during the acceleration to this condition and second, during the initial part of the coast phase after thrust cut off. In the first case, abort before injection, the requirement for abort propulsion rapidly decreases with decrease in velocity. This reduction occurs because the flight path angle and altitude are less severe and also because, at the reduced velocity, the skip out boundary angle is less negative than at escape speed. Thus, a system capable of safe abort at injection has excess propulsion for abort at any time up to injection. This extra propulsion may be used to provide improved navigational flexibility for reaching a desirable landing point after entry.

For abort after injection the propulsion requirement increases to the same extent shown on Figure 7 for the lower injection velocity conditions. Thus, provision of abort capability to 1,200,000 ft on the Earth-Moon trajectory would require at that altitude the same propulsion required for abort at injection to this condition. This large requirement may indicate that abort appreciably after injection is prohibitively expensive particularly if it is considered that the highest likelihood of failure requiring abort occurs during the thrusting and cut-off phases.

If a weight analysis similar to that shown on Figure 8 was made utilizing an S-IVB stage modified to increase the thrust-to-weight ratio to near 1, it is apparent that the drop off in performance associated with injection near perigee would not occur. Thus the optimum injection conditions would be at perigee if no limitation on abort propulsion were imposed. Or, if as in the Apollo, the abort propulsion were used for a mission requirement the optimum condition would be that at which the mission propulsion was sufficient to provide safe abort. If in this case the optimum injection conditions were near perigee, an appreciable performance gain over the low T/W S-IVB stage could be obtained. The increased performance potential would, of course, have to be balanced against the increased inert weight of the higher thrust system.

CONCLUDING REMARKS

A brief analysis of the interactions between abort considerations and the selection of injection conditions to a representative Earth-Moon trajectory has been performed in order to provide a quantitative evaluation of the design and operational choices which maximize the mission effectiveness. It has been found that the optimum combination of these effects is dependent on the design characteristics and requirements of the system. A procedure for the evaluation of these interactions for representative systems is presented and its use illustrated by an example.

The analysis has resulted in the following general conclusions:

1. The payload ratio from Earth orbit to Earth-Moon injection is maximized by use of appreciable positive injection angles.

2. Injection into an Earth-Moon trajectory at or near perigee compared to injection at the optimum point results in a payload penalty which is relatively insignificant for initial thrust-to-weight ratios of one; the penalty becomes large at initial thrust-to-weight ratios of the order of one-half. Injection payload is maximized by use of an Earth-Moon trajectory with perigee above the orbit altitude.
3. The penalty for the provision of abort to immediate Earth atmosphere reentry increases from a moderate value for injection at perigee to a very large value for injection at the optimum point from performance considerations alone. The abort penalty increases rapidly with reduction of the thrust-to-initial-weight ratio of the abort propulsion system.
4. For an example Apollo system, the optimum injection condition considering both payload and abort requirements occurs near perigee. The use of a higher thrust-to-initial-weight ratio in the S-IVB stage would provide a considerable increase in the flexibility of operation of the system and would probably result in some increase in useful payload.

REFERENCES

1. Anonymous, "Dynasoar Phase Alpha Comparative Configuration Study," (U) Lockheed Aircraft Corporation, California Division, Burbank, Report No. 14295, February 22, 1960. (Secret)

2. McGowan, W. A. and Eggleston, J. M., "Preliminary Study of the Finite Thrust Engines for Abort During Launch of Space Vehicles, NASA TN D-713, December 1961.

3. Anonymous, "Preliminary Analysis of the Fuel Required for Abort at Earth Escape Velocity," NASA, Apollo working paper No. 1008, February 2, 1961.

4. Anonymous, "Proposal for project Apollo Feasibility Study," Lockheed Missiles and Space Division. Report 288700, October 7, 1960.

5. Huss, C. R., Hamer, H. A. and Mayer, J. P., "Parameter Study of Injection for Lunar Mission Including Various Trajectory Considerations, NASA TR - 122, 1961.

A PROCEDURE FOR CALCULATING RADIATION EXPOSURE ON SPACE MISSIONS

R. A. Miller
W. Cranford

General Dynamics/Fort Worth
Nuclear Aerospace Research Facility
Fort Worth, Texas

ABSTRACT

A problem associated with manned space missions is the determination of shielding requirements for the protection of man from the hazards of space radiation. A space trajectory radiation exposure has been formulated for estimating the magnitude of this radiation for missions in cislunar space. The results are described for some calculations of the dose incident on a vehicle subjected to trapped radiation, cosmic radiation, and a solar-flare radiation during a lunar mission of about seven days.

The work described in this paper was performed under Air Force Contract No. AF33(657)-7201.

A PROCEDURE FOR CALCULATING RADIATION EXPOSURE ON SPACE MISSIONS

R. A. Miller and W. Cranford
General Dynamics/Fort Worth
Nuclear Aerospace Research Facility
Fort Worth, Texas

INTRODUCTION

A major goal of the space program is the exploration of the planets of the solar system. In particular, efforts are being concentrated toward early exploration of the moon, which is the nearest and most promising celestial body for early manned missions. The problem described in this paper is concerned with one phase in determining the requirements for the protection of man from the hazards of space radiation. A computer program, STREP, has been developed at General Dynamics/Fort Worth under Air Force sponsorship to determine the magnitude of the radiation hazard to which manned space vehicles are exposed while on missions in cislunar or lunar space. The results of this program can be used as input for another procedure to calculate the dose inside a shielded crew compartment. The results of calculations made with STREP for two lunar missions, one in which the vehicle is beyond the earth's magnetic field during a solar flare and the other in which the vehicle is in a 24-hour orbit during the solar activity period, are described. Some implications are indicated for the secondary components of the dose within a shielded crew compartment.

HAZARDS TO SPACE MISSIONS

The discovery of ionizing radiations surrounding the earth has emphasized the existence of a hazard which must be assessed before manned space flight is feasible. From balloon observations and satellite measurements, it is evident that the intensities of charged particles, both protons and electrons, are sufficiently great to create a radiation hazard. Detailed investigations are necessary to evaluate the data and to calculate the magnitude of the radiation so that adequate protection, e.g., shielding, may be provided for manned space missions.

The penetrating radiations to which a manned space vehicle will be exposed may be divided into four broad classifications: primary galactic cosmic radiation, Van Allen radiation, solar-flare radiation, and miscellaneous extraterrestrial radiation.

Of these categories, the greatest hazard to crews on lunar missions will be due to the solar-flare protons.

Cislunar space is divided into five domains by virtue of the geomagnetic field and the incident solar and primary cosmic particle radiations. Each of these regions may influence manned space operations and may be described briefly as follows. The first zone, which is restricted to low geomagnetic latitudes and altitudes, has a paucity of trapped particles and is relatively safe for manned flights without shielding. In the second zone, trapped protons with peak intensity near 3600 kilometers in altitude present a major hazard to man and necessitate shielding for trajectories requiring extensive operation within the zone. In the third zone, which extends beyond the second zone and varies in size and intensity of the trapped particles during solar activity, the principal hazard arises from electrons of such intensity that a bremsstrahlung problem may be present. In the fourth zone, over the magnetic poles, the proton flux from solar flares is of such magnitude that unshielded manned vehicles are unsafe at altitudes above the earth's atmosphere. In the last region, which is the transgeomagnetic space in which the earth's magnetosphere does not modify the trajectories of charged particles, the principle hazard is from solar-flare protons. In this case, radiation exposure to primary protons and secondary protons and neutrons produced in the vehicle may exceed acceptable levels for the protection of man without shielding. Additional potential radiation hazards may be associated with primary cosmic radiation, heavy particles, and secondaries originating in the atmosphere.

For the purposes of the discussion in this paper, a brief description of the present estimates of the intensities and spectra for the radiation in space will suffice. The early conceptual model of the spatial distribution of the Van Allen radiation consisted of two concentric toroidal regions extending equatorially about the earth. For this model the inner zone extended from about 600 kilometers to about 10,000 kilometers in altitude between geomagnetic latitudes of 25°S and 25°N.

Van Allen (1) has estimated that the omnidirectional intensity for protons of energy greater than 40 Mev is on the order of 2×10^4 protons/cm^2-sec in the heart of the belt. Figure 1 shows inner-belt proton energy spectrum derived by Freden and White (2). In Figure 2 is a plot in geomagnetic coordinates of the proton isocount contours normalized to the Van Allen value of the proton intensity.

Holly, et al. (3) have published data of measurements of the inner-belt electron spectrum. Estimates of the maximum unidirectional intensity for electrons of energy greater than 20 kev are given by Van Allen (1) to be on the order of 2×10^9 particles/cm^2-sec ster.

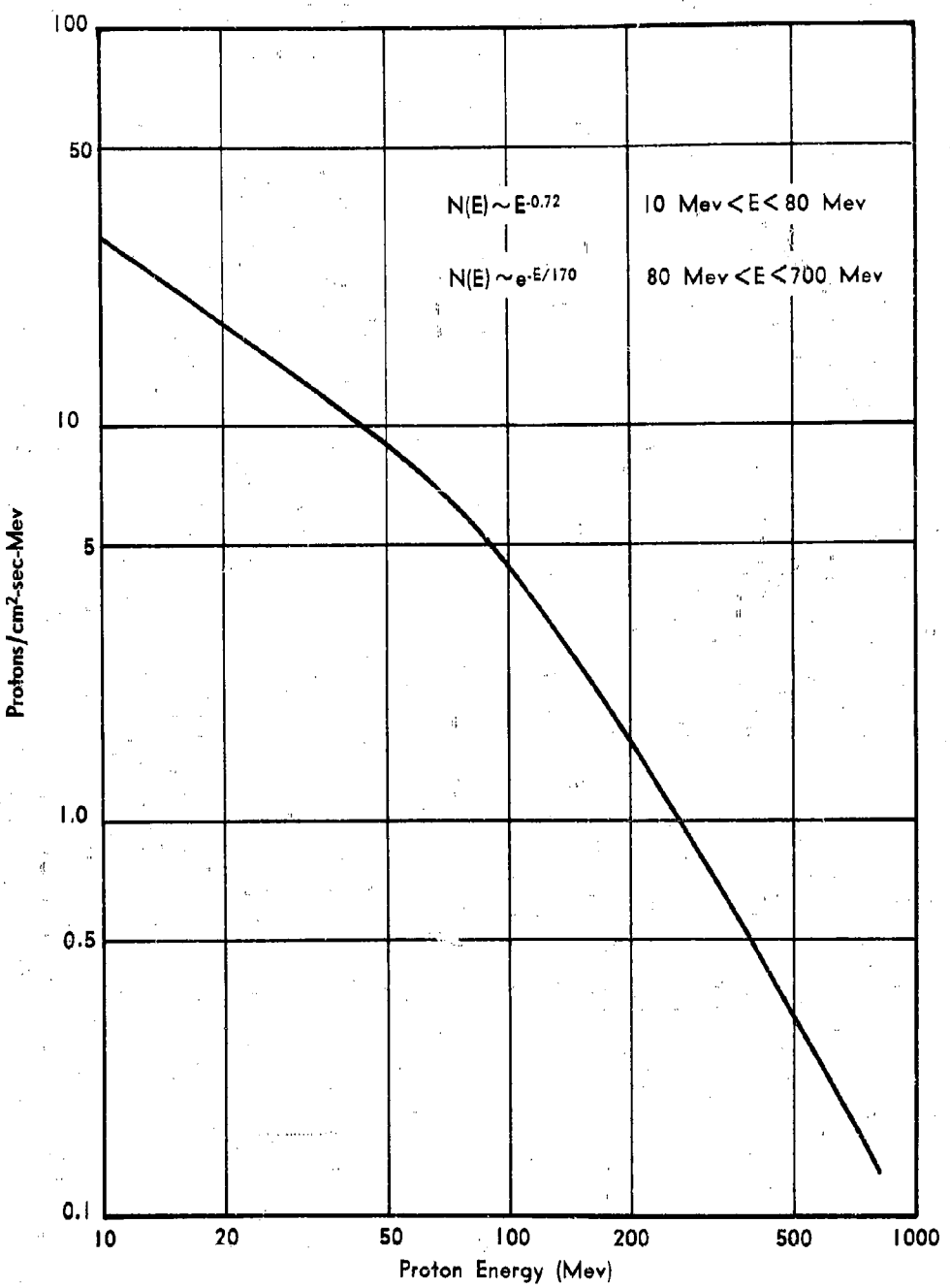


Fig. 1. Inner-Belt Proton Spectrum (Freden and White)

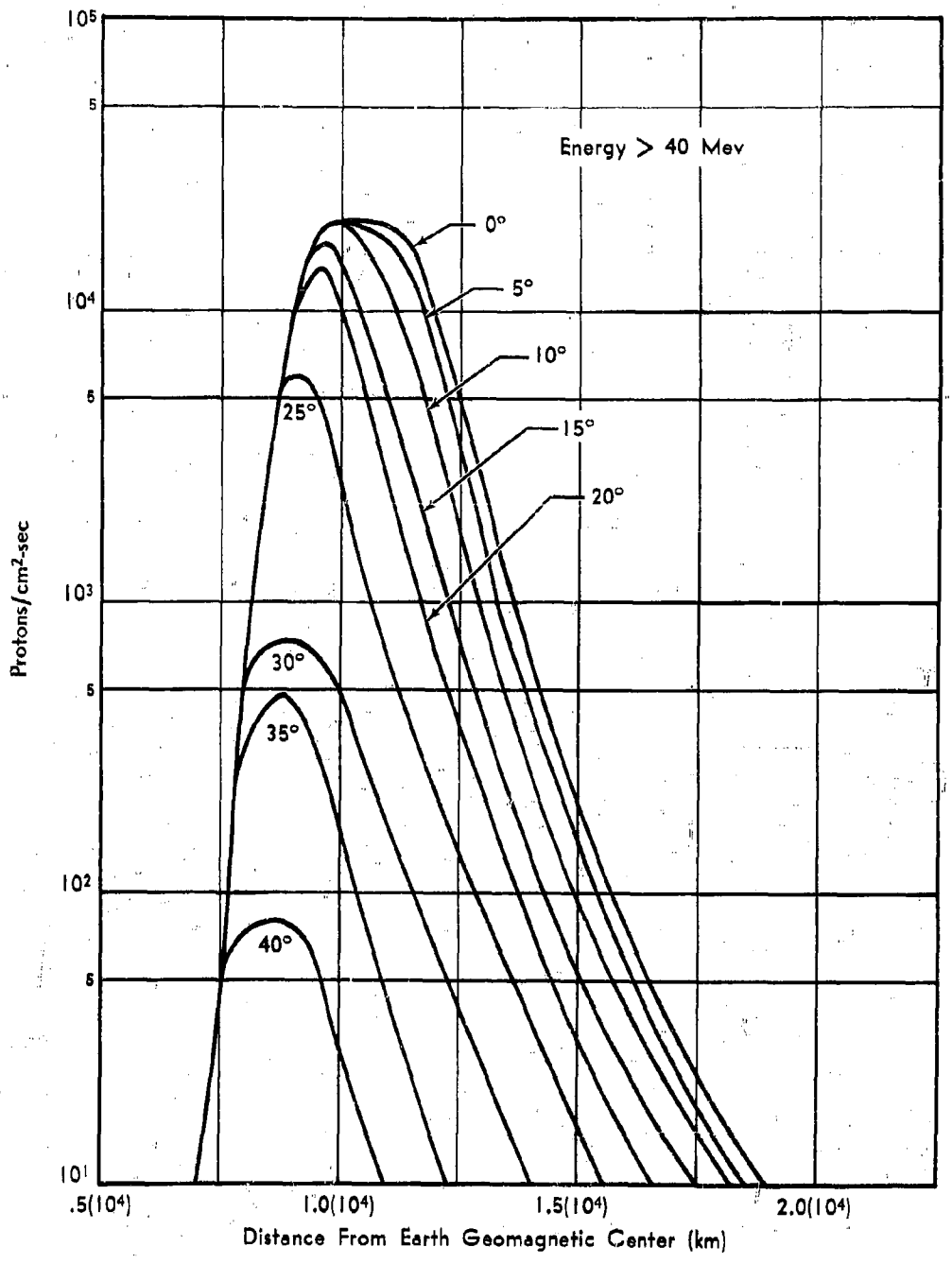


Fig. 2. Inner-Belt Proton Flux

The omnidirectional electron flux in the outer belt has been the subject of considerable disagreement. The uncertainty stems from the inherent error involved in measuring electrons unable to penetrate the shielded instruments. Some flux contours are shown in Figure 3 for electrons in the radiation belts for both quiet and excited days. The excited-day values are for periods immediately following magnetic storms.

Solar flares in general have an extreme time variability, which may depend to some extent on sunspot activity, since the flare frequency apparently varies roughly as the sunspot number with a period of approximately 11 years. During the peak sunspot cycle, Class 3+ flares may average one or more per month; however, during July 1959, three Class 3+ flares were observed within six days. The Class 3+ flares can be considered as two types: high-energy and low-energy events. During the high-energy events, protons are observed arriving at the earth with relativistic velocities and with energy estimates up to 20 Bev.

Low-energy (non-relativistic) solar-flare events consist mainly of protons in the energy range between 40 and 500 Mev; thus the energy is low enough that the particles are not observed at sea level because of the magnetic cutoff.

The energy spectrum for the 14 July 1959 solar flare, measured 32 hours after onset as reported by Winckler et al. (4), is shown in Figure 4. Data for this flare used in the example described later in this paper are listed in Table 1.

Table 1. Solar-Flare Data

Date	Differential Energy Spectrum (proton/cm^2-sec)	Energy Range (Mev)	Decay (hr)	Total Time After Flare (hr)
14 July 1959	$1.5 \times 10^7 E^{-3.9}$	$30 < E < 1000$	t^{-2}	55.5

Although the major radiation component from solar flares consists of protons, some evidence of solar gamma radiation, heavy particles, and α-particles has been noted. For example, from the observations on one rocket flight, during the 12 November 1960 flare, Biswas, et al. (5) give intensities in the energy interval 42.5 to 90 Mev/nucleon for α-particles and heavier nuclei as approximately 12 and 0.2 particles/cm^2-sec-ster, respectively. In view of the low energy of these particles, they may be totally absorbed in the shielding thickness necessary to attenuate the proton component to an acceptable dose level.

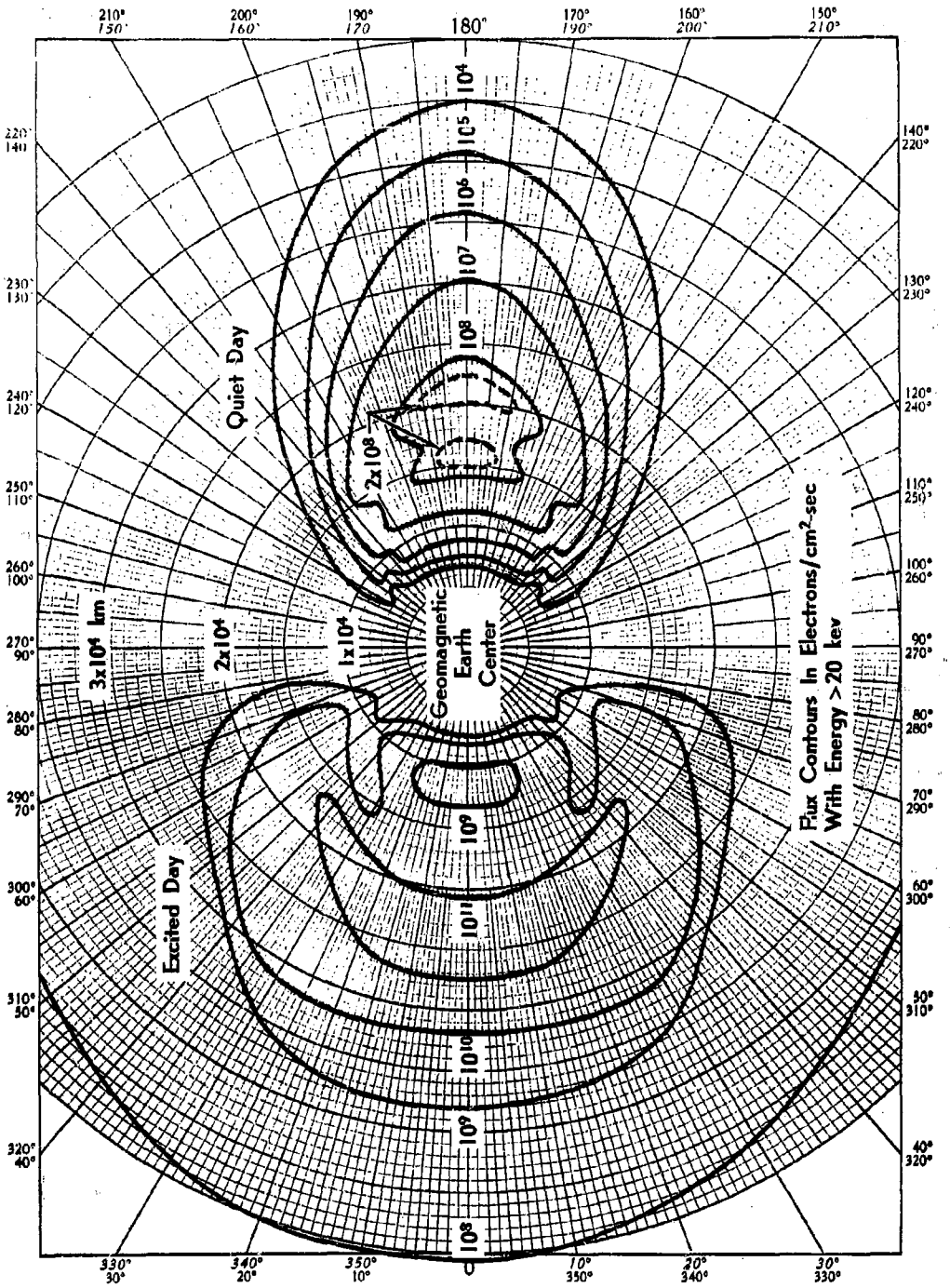


Fig. 3. Electron Belts

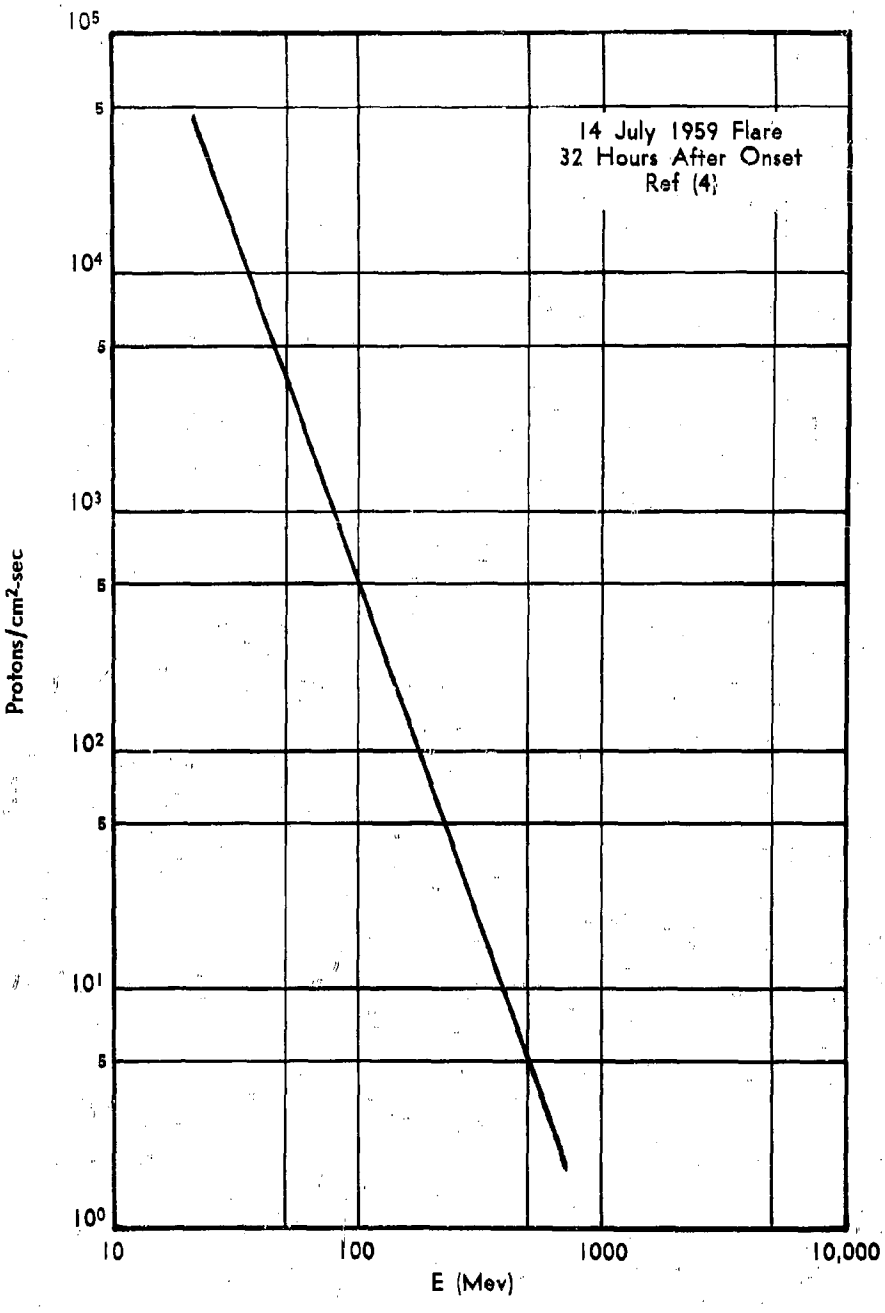


Fig. 4. Solar Flare Proton Energy Spectrum

In the arrival near the earth of charged particles from the sun, the high-energy particles are observed arriving ahead of the low-energy particles. Subsequent to the onset of the flare, a directional flux of high-energy particles has been observed to last from 10 minutes to several hours (6), followed by an isotropic flux of lower-energy particles lasting several hours. Studies on the energetic prompt events indicate that the particles appear to come from a source about 40° - 55° west of the sun. The angular distribution from the solar flare of 28 September 1961 reported by Van Allen (private communication) was markedly anisotropi as measured on Injun I at the 1000-kilometer altitude. The isotropy of the greater portion of the solar flux is presumably due to the operation of a storage mechanism in interplanetary space.

Additional potential radiation hazards may be associated with primary cosmic rays and certain secondaries originating in the atmosphere.

CALCULATION OF RADIATION EXPOSURE

A method has been developed for determining the shielding requirements for the protection of the space-vehicle crews. An existing computer procedure which calculates the time-integrated spectra for mission orbits using available radiation data will be described in the sections that follow. These spectra can be used in a second computer program, Proton Attenuation Procedure (PAP), to calculate the primary and secondary proton and secondary neutron dose components within a shielded vehicle.

The Space Trajectory Radiation Exposure Procedure (STREP) was designed to compute the time-integrated spectra for any desired trajectory in cislunar or lunar space and for any combination of the several components of space radiations. These components include the following: Van Allen protons and electrons; solar-flare protons, gammas, electrons, and heavy particles; cosmic-radiation protons and heavy particles; albedo neutrons; and aurora borealis gammas. The program will also calculate the dose behind a thin shield at any time after the start of the mission. For shielded vehicles, the time-integrated spectra may be used as input for the more sophisticated penetration procedure (PAP) to secure the dose as a function of shield thickness.

THE TRAJECTORY CALCULATIONS

Integration of the space radiation spectra over a complex vehicle mission using experimental data for the radiation fields requires numerical integration over time. This requirement, in turn, implies the knowledge of the vehicle position as a function of time. The number and spacing of this set in space and time necessary to obtain the desired accuracy, under the assumption of accurate knowledge of the radiation field, influence the

general approach to the trajectory computation problem. The spatial separation of the required points is governed by the gradients of the radiation fields.

Since the problem has been confined to cislunar space and no spatial dependence of the radiation components is known to exist in the vicinity of the moon, consideration is given only to the gradients induced by the earth. Such gradients are known to exist out to 7-10 earth radii, thereby defining the sphere of influence corresponding to the earth's magnetosphere. Outside this sphere of influence in cislunar space, all types of radiation are assumed to be independent of position; however, time variations of components such as solar flares will influence the integration. The greatest gradients are those associated with the Van Allen radiation regions. In practice, short time steps in integration are required inside the sphere of influence, whereas long steps may be used outside this region.

Since most of the space missions of immediate interest involve either close approach to an astronomical body other than the earth or several earth orbits, more than two points on the trajectory are necessary as input information for two-body calculations. For several reasons, a decision was made to obtain the perturbation results from an external source rather than attempt to generate the necessary set of points within the procedure for the two-body calculations from the desired characteristics of the trajectory. Accordingly, the necessary portions of two codes, developed at General Dynamics/Astronautics (7), have been incorporated in an interpolation subroutine of STREP. It is assumed that a list of positions and corresponding times is available within restrictions on two-body orbits.

In order to reduce the number of positions required for input, the procedure solves the two-body problem by using an initial position vector, estimated initial velocity vector, and the elapsed time to reach a second position. It computes the free-fall orbit necessary to reach the second position at the proper time. Once these parameters are established, STREP can compute any position along this trajectory leg needed by the integration procedure. For free-fall orbits inside the earth's magnetosphere, positions can be given at time intervals of less than one period. However, for thrust portions of the trajectory inside the sphere of influence, positions must be given every thirty seconds to maintain accurate integration. Outside the sphere of influence for either thrust or free-fall legs, the positions must be given only when a change from thrust to free-fall, or vice versa, is evident. At each initial position of a free-fall orbit leg, an estimate of the velocity at this position must be given unless the previous leg was free-fall. The time from the start of the mission must be specified with each position. Any number of thrust and free-fall legs may be used to

describe the trajectory, with the necessary position data given in any one of several different coordinate systems.

For a trajectory for which burnout occurs below the inner Van Allen belt, calculations are begun with burnout as the first point. The first two points (P_1, P_2) and the elapsed time are used to compute a required velocity, with an estimate of burnout velocity and an initial guess for the flight-path angle. The radiation intensities are evaluated at the first point, P_1. A time step is calculated from the initial point by the equation

$$\Delta t_1 = K\, r^{3/2}, \qquad (1)$$

where r is in earth radii and K is an empirically determined constant. The radiation integration is performed with the integrand as a product of the intensity at the initial point and the time interval. The position at the end of the time interval is computed from the initial point, with the required velocity and radiation intensities evaluated at the end of the first step. After a new time interval is found from the new position, the radiation integrals are increased by the amount of the contribution of the second step. The process is continued, with the new positions, P_{1i}, determined by the sum of the time intervals from the initial points until this sum equals or exceeds the time between P_1 and the second input point P_2. The velocity at the end of this first leg is computed and is used as the initial guess for the second leg. This process is continued until the vehicle leaves the magnetosphere or else enters a high-thrust leg condition, which is indicated by an input control number. In either of these cases, the two-body interpolations are not used until the vehicle is back within the magnetosphere or is on a low-thrust leg.

In Figure 5 a schematic diagram is shown of the trajectory characteristics as treated within the program. The portions of the trajectory, high-thrust and external to the magnetosphere in which interpolations are not carried out, are distinguished from the free-fall legs inside the magnetosphere. An indication is given for the type of radiation to which the vehicle is exposed during the mission.

RADIATION CALCULATIONS

In addition to the positions along the trajectory, data must be given for each type of radiation expected on the mission. Although STREP does not predict the time of occurrence of solar flares or magnetic storms, any combination of these events with the associated times from the start of the mission may be used in describing the radiation environment. A maximum of ten changes, e.g., a total of ten different solar flares, could be provided for each mission.

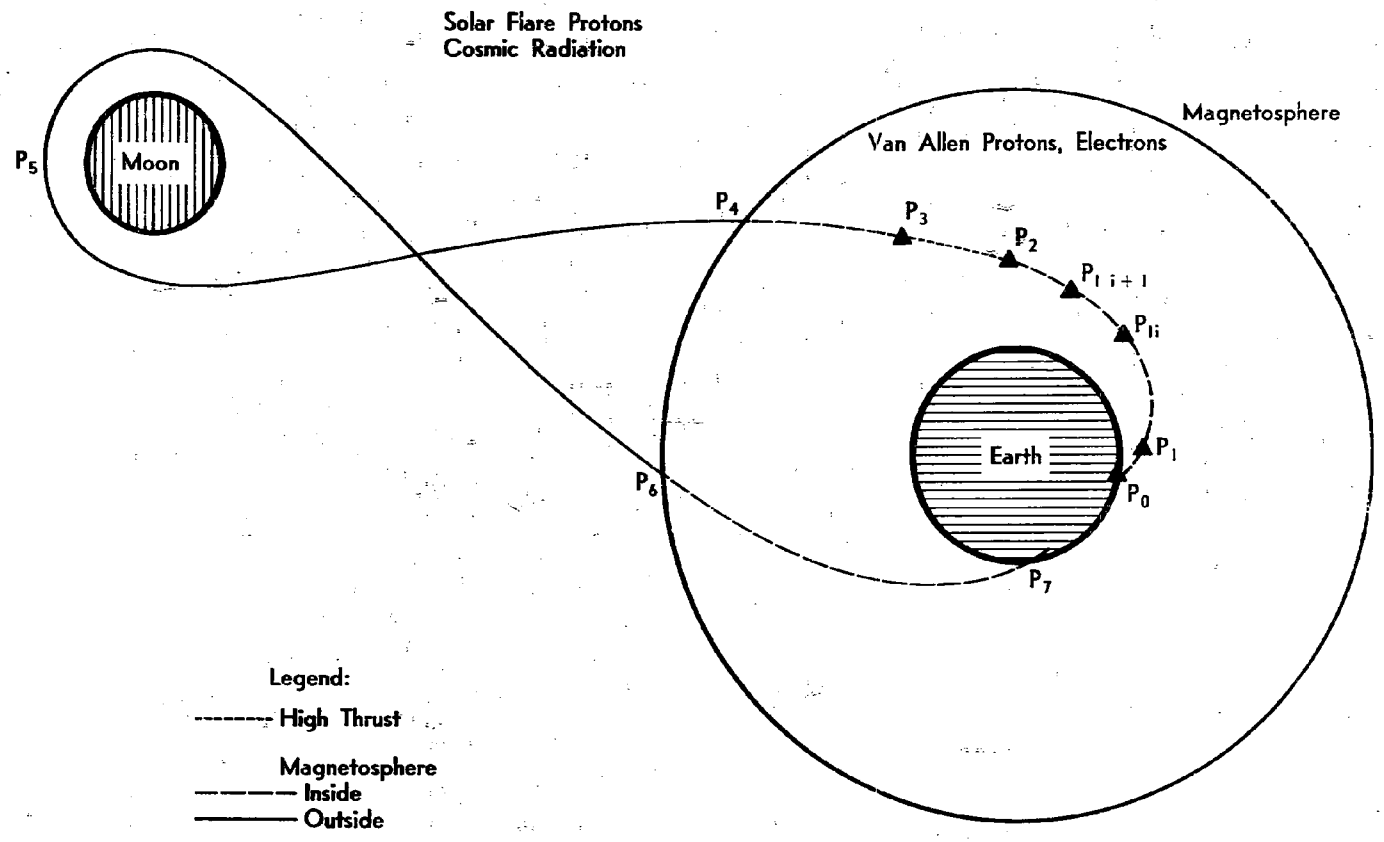


Fig. 5. Trajectory Characteristics

A brief description of the computational method for each type of radiation component is given in the following paragraphs. For Van Allen protons and electrons the intensity is secured by double interpolation in a two-dimensional space-dependent array. The spectrum is computed from an expression of the form

$$\psi\ (E) = N_1\ e^{-k_1 E} + N_2\ E^{-k_2}, \qquad (2)$$

where the four constants N_1, N_2, k_1 and k_2 are space dependent. Provision is made to expand and contract the Van Allen belts in case of a magnetic storm by insertion of new data at the proper time. A numerical integration is performed to secure the time integrated spectra.

The spectra for solar-flare electrons, protons, and heavy particles are all computed by a formula similar to $\psi(E)$, above, with a decay factor t^{-k} and a normalization constant secured by an integration over the energy range. Outside the earth's magnetosphere the spectra are integrated analytically, whereas inside the earth's magnetic field the spectra must be modified and integrated numerically because of the effect of the field on charged-particle intensities.

In regard to charged particles inside the earth's magnetosphere, an approximation is acquired through the reduction of the intensity at each energy by the solid-angle effect imposed by the cutoff energy and the distance from the earth's center.

For cosmic radiation, the spectra are computed by the formula

$$\phi(E) = C_1\ (C_2 + E)^{-k}, \qquad (3)$$

which may be overridden by input spectra data. These spectra are assumed independent of position for points outside the influence of the earth's magnetic field. When inside this field, the calculation is modified by consideration of the effect of the field on charged particles. The program represents the range of heavy particles by types such that the parameters in $\phi(E)$ are dependent on the type of particle.

Figure 6 shows a flow diagram of the computation within the procedure.

RADIATION CALCULATION FOR A LUNAR MISSION

To provide some quantitative evidence of the radiation to which an unshielded manned space vehicle on a lunar mission would be exposed, two flights were programmed on STREP. A trajectory

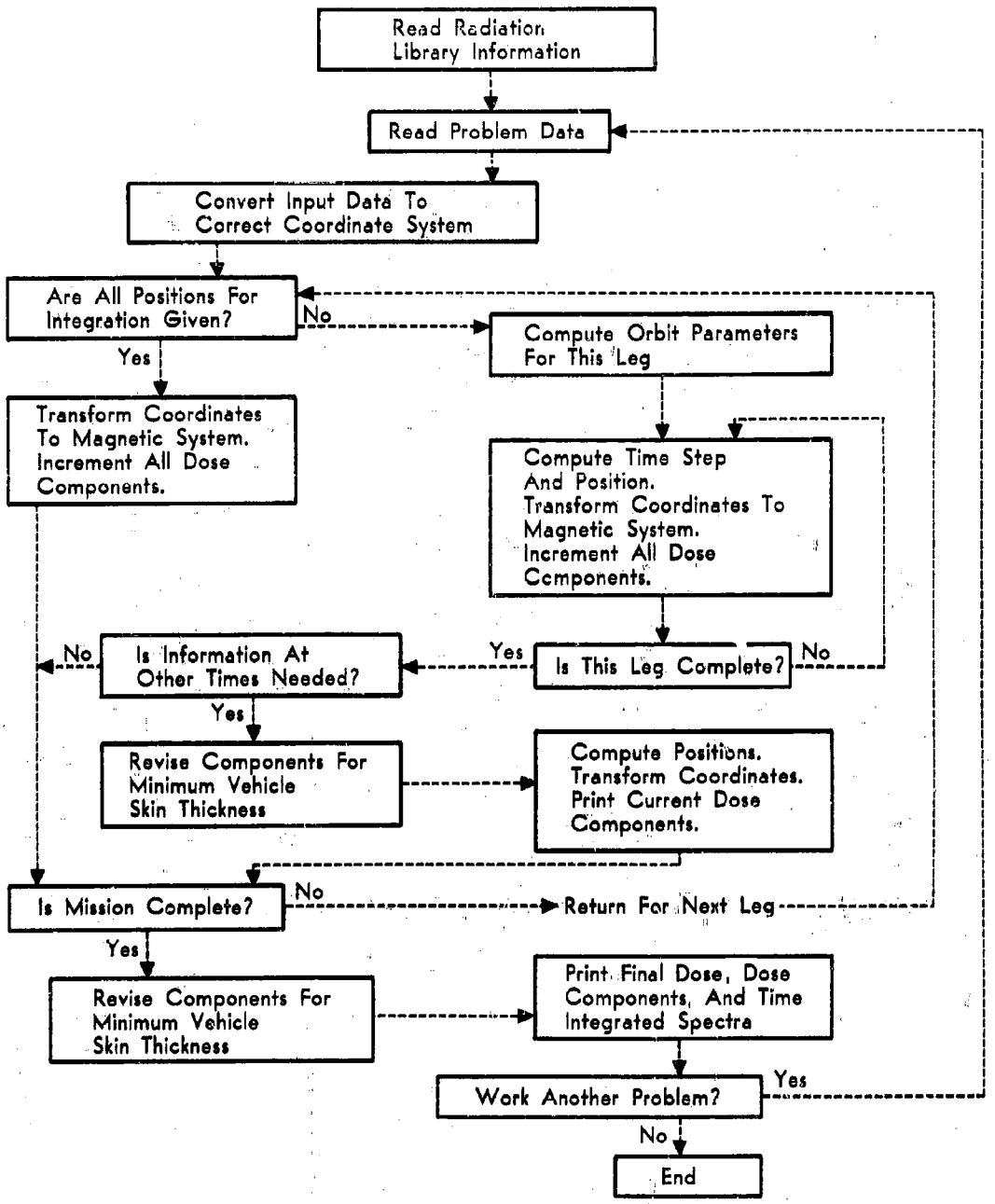


Fig. 6. STREP Functional Chart

Table 2. Trajectory Specifications

Conditions	Trajectory 1	Trajectory 2
Launch		
Latitude	28.278°N	Same as Traj. 1
Longitude	80.574°N	
Heading Angle	98.88°	
Time to injection	785.0 sec	
Injection		
Time	20 January 1968 2255UT	Same as Traj. 1
Latitude	17.25°N	
Longitude	45.35°W	
Heading Angle	114.34°	
Altitude	54.43 miles	
Velocity	36,167 ft/sec	
Radiation Exposure (hours after injection)		
Van Allen Proton Belt (exit)	.82	.82
Van Allen Electron Belt (exit)	3.4	3.4
Circular orbit (enter)		10
Solar Flare (onset)	12	12
Solar Flare (end)	67.5	67.5
Circular orbit (end)		70
Re-entry		
Time from injection	149.2 hours	209.2 hours
Altitude	75.77 miles	75.77 miles
Longitude	154.6°W	154.6°W
Latitude	6.4°N	6.4°N
Velocity	36,082 ft/sec	36,082 ft/sec

of 149.2 hours duration from the point of injection to the point of re-entry (8) was used as a model.

In the first trajectory the vehicle was permitted to traverse the Van Allen regions, to proceed about the moon while exposed to a solar flare of the 14 July 1959 type, and then to return through the belts to the re-entry point. In this case the flare was of 55.5 hours duration and was timed to begin onset at 12 hours after injection of the vehicle into its trajectory.

In the second case, the vehicle traversed the belts, proceeded beyond the magnetosphere, and received warning of probable flare activity six hours prior to onset. The vehicle returned to circle the earth in a 24-hour orbit (22,400-mile altitude) for 60 hours and then completed the lunar mission. A schematic drawing of two trajectories is shown in Figure 7. The trajectory specifications and the data on the solar-flare events are listed in Table 2 (8). The trajectory data taken from (8) and shown in Table 3 were used as input for the calculations in STREP. Some indications of the accuracy of the STREP calculations of the required velocity can be seen from a comparison of the results in Table 4.

Table 3. Trajectory Data

Point	Time (hr)	Alt. (mi)	Vel. (ft/sec)	Lat.	Long.
1	0	54.4	36,167	17.25°N	45.35°W
2	.2	1,380	31,281	10° S	2° E
3	.9	9,548	19,366	27° S	46° E
4	3.9	34,934	10,838	29° S	36° E
5	7.9	58,886	8,077	26° S	15° W
6	136.1	83,652	6,485	29° S	114° W
7	145.9	30,566	11,644	39° S	115° E
8	149.0	1,149	32,003	25° S	177° E
9	149.2	75.77	36,082	6° N	155° W

Table 4. Velocity Comparisons

Time (hr)	Velocity (ft/sec)	
	Table 3	STREP
0	36,167	35,805
7.9	8,077	8,109
136.1	6,485	6,478
149.2	36,082	36,071

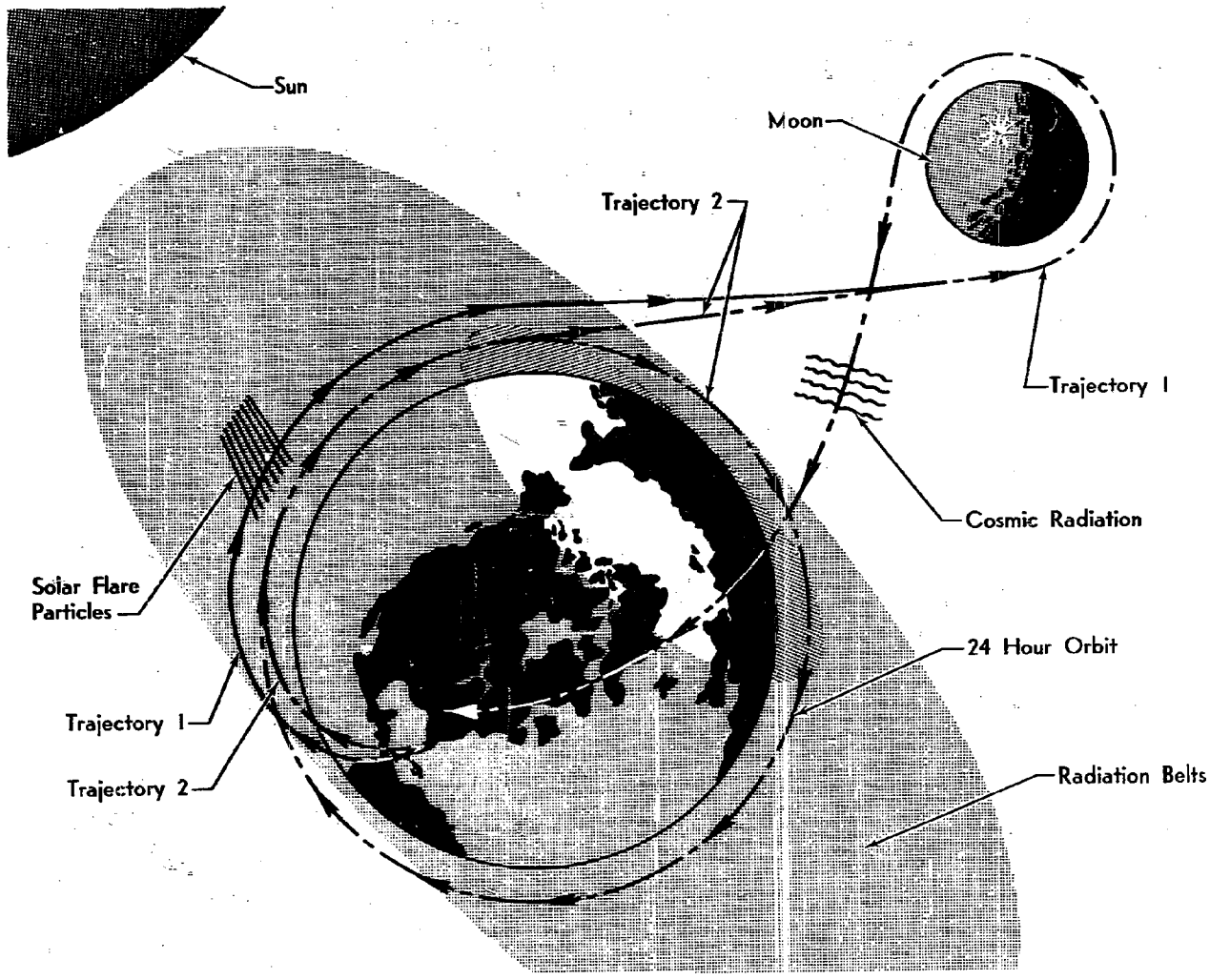


Fig. 7. Lunar Trajectories

The time-integrated spectra incident on the unshielded vehicle from the solar flare protons, Van Allen protons, and cosmic radiation for the trajectories are shown in Figure 8. It will be noted that the solar-flare spectrum for Trajectory 2 is lower than that for Trajectory 1 in the energy range of from 30 to 300 Mev. This decrease for the lower-energy particles will reflect a decrease in the dose arising from the secondary particles born within a shield material.

Figure 9 shows the time-integrated spectra for three different types of heavy particles for Trajectory 1. A comparison of the time-integrated spectra for electrons during the two trajectories is shown in Figure 10. A slight increase of the electron spectrum, the heavy-particle spectrum, and cosmic-radiation spectrum was calculated for the 60 hours spent in the 24-hour orbit.

The results of the unshielded-dose calculations behind an aluminum skin of 2 gm/cm^2 in thickness are summarized in Table 5 for the two trajectories mentioned above and for Trajectory 3, which is a modified Trajectory 1. This modification amounted to a change of 30° in the longitude of the end point of the first leg of the trajectory after injection. The original position (27°S, 140°W) was changed to (27°S, 16°E). A portion of Trajectory 1 and Trajectory 3 through the inner Van Allen proton belt is shown in Figure 11. This judiciously chosen alternate trajectory essentially avoided the radiation belts, with a resulting decrease in the total Van Allen dose from 14.3 rads to 2.9 rads. It may be remarked that an equatorial orbit selected at about 2-1/2 earth radii would have eliminated the contributions from the flare.

Table 5. Unshielded-Dose Calculations

Component	Dose (rads)		
	Trajectory 1	Trajectory 2	Trajectory 3
Van Allen Protons	$3.1(10^{-1})$	$3.1(10^{-1})$	$6.8(10^{-3})$
Solar-Flare Protons	$2.2(10^{4})$	$9.4(10^{3})$	$2.2(10^{4})$
Cosmic Protons	$6.2(10^{-2})$	$8.2(10^{-2})$	$6.2(10^{-2})$
Heavy Particle A	$2.7(10^{-2})$	$3.8(10^{-2})$	$2.7(10^{-2})$
Heavy Particle B	$7.9(10^{-3})$	$1.1(10^{-2})$	$7.9(10^{-3})$
Heavy Particle C	$2.5(10^{-5})$	$3.5(10^{-5})$	$2.5(10^{-5})$
Van Allen Electrons	$1.4(10^{1})$	$1.7(10^{1})$	$2.9(10^{0})$
Albedo Neutrons	$1.3(10^{-4})$	$2.7(10^{-4})$	$1.8(10^{-4})$

It must be understood that the above dose values are for an essentially unshielded vehicle. The results of using the

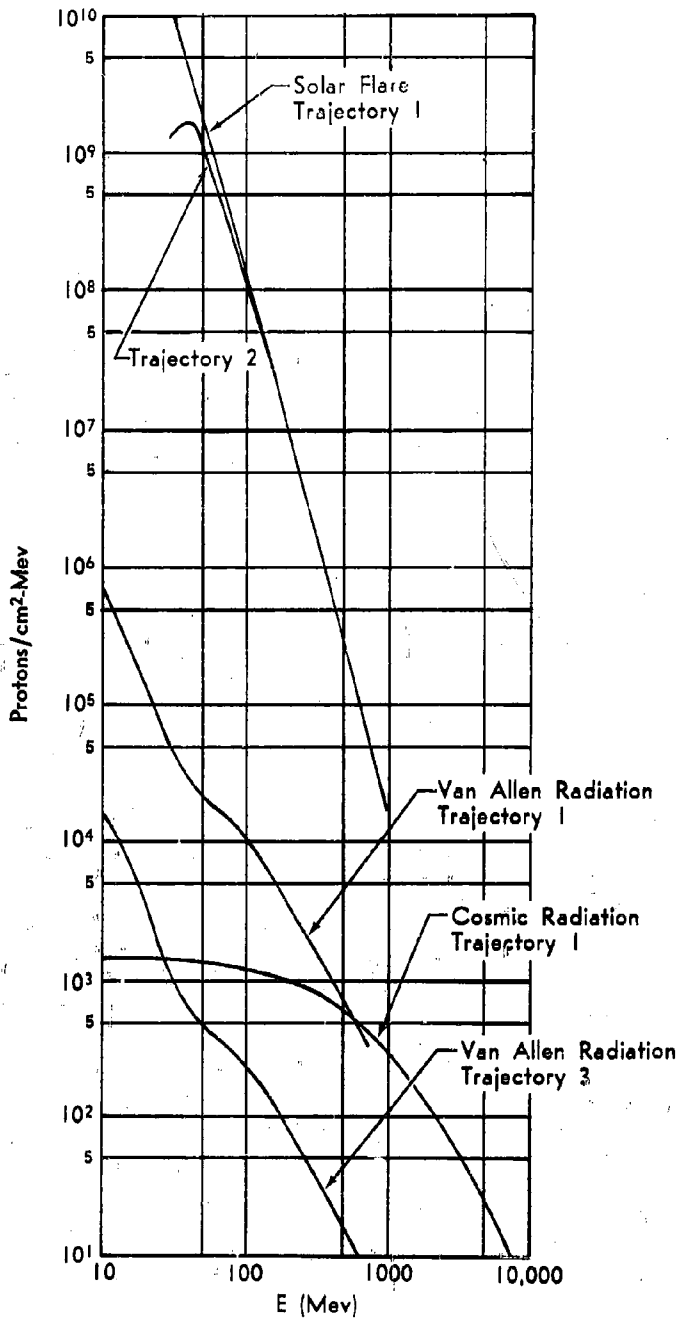


Fig. 8. Proton Time-Integrated Spectra

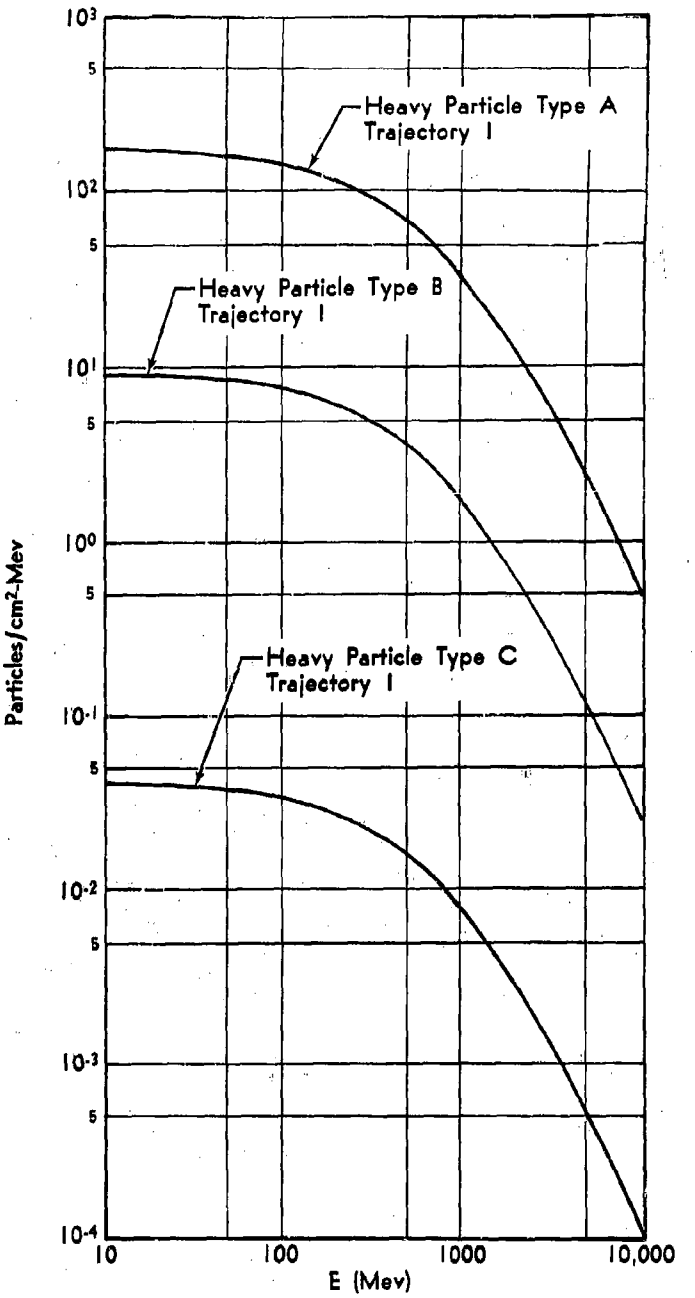


Fig. 9. Heavy-Particle Time-Integrated Spectra

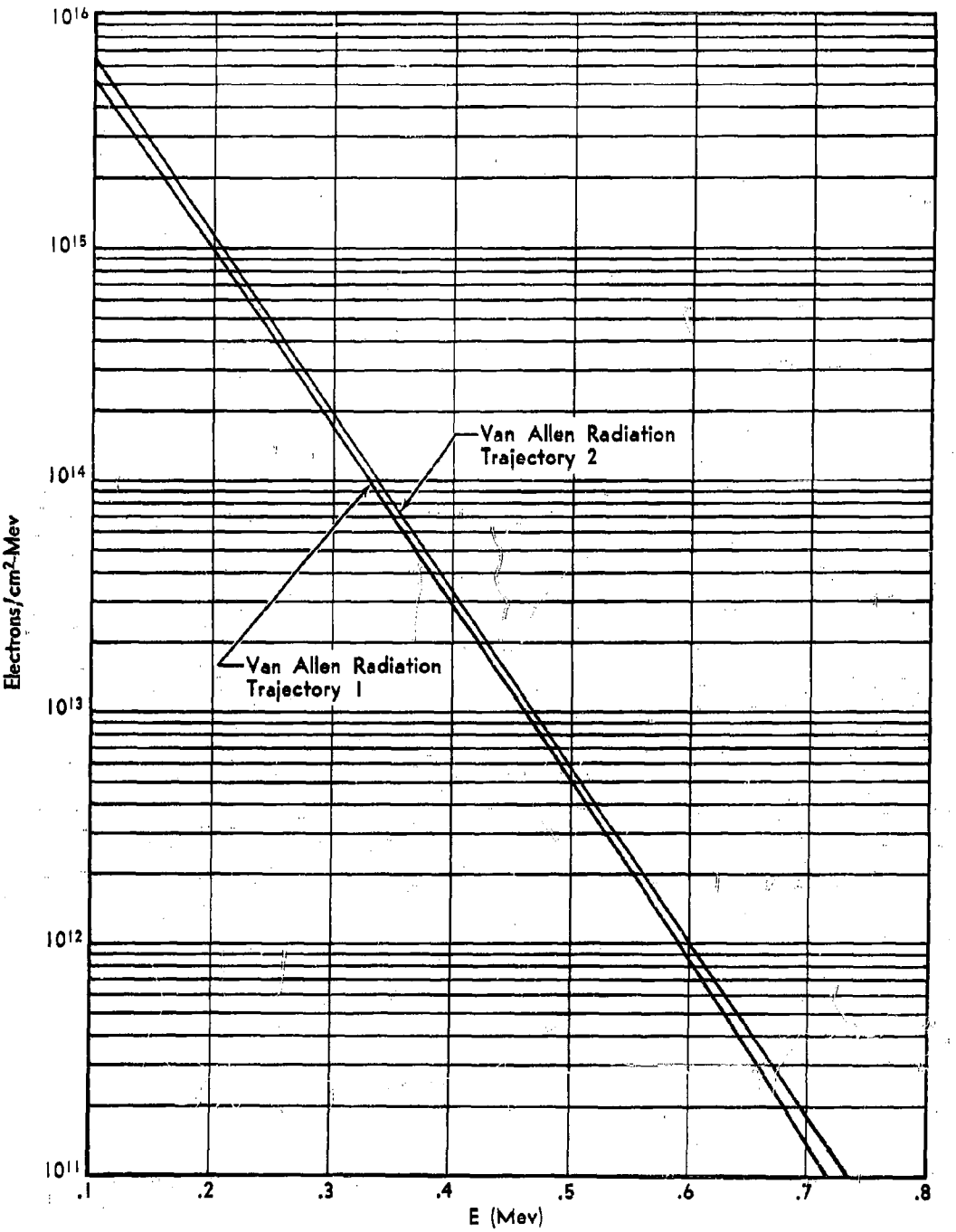


Fig. 10. Electron Time-Integrated Spectrum

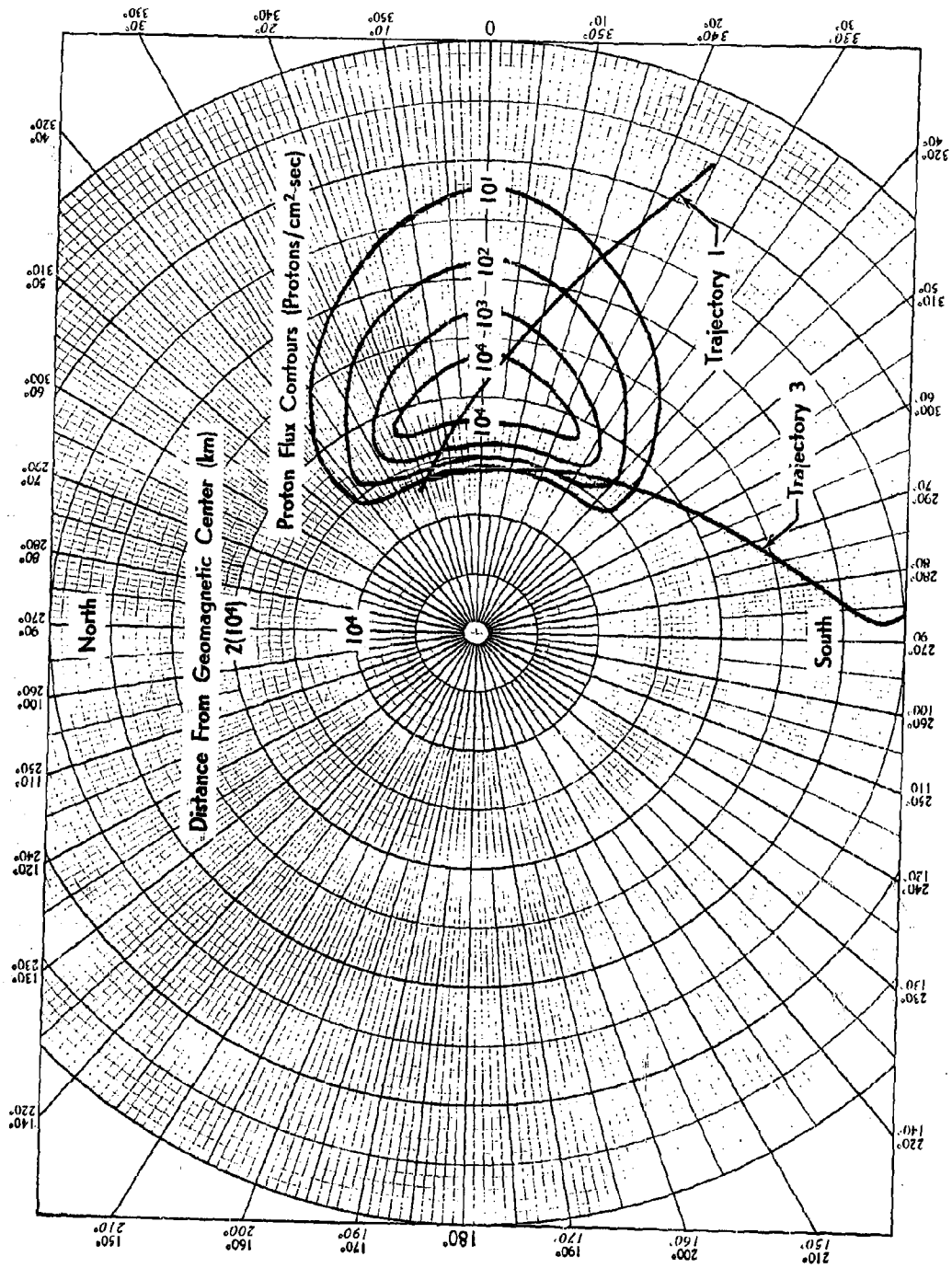


Fig. 11. Trajectories Through Van Allen Proton Belt

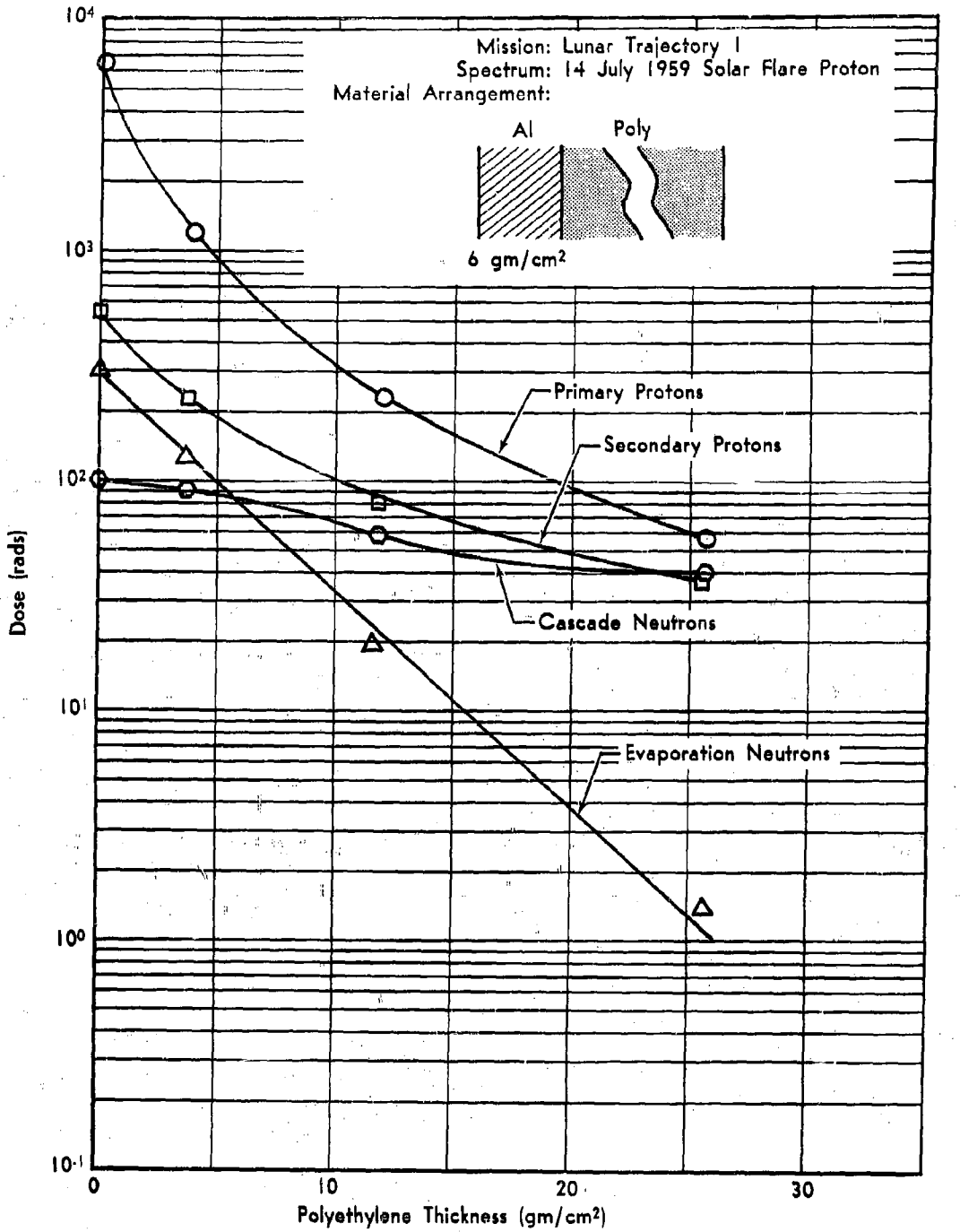


Fig. 12. Composite-Shield (Aluminum-Polyethylene) Dose as Function of Polyethylene Thickness

spectra output in PAP are shown in Figure 12 for the lunar Trajectory 1. A fixed shield of 6 gm/cm^2 of aluminum backed with a variable thickness of polyethylene was assumed. The several components of the dose, primary and secondary, are shown as functions of the polyethylene thickness. It should be noted that the contribution to the dose from the secondary components begins to approach the primary component with increasing thickness of the shield. This predominantly secondary component has been characteristic of other calculations of dose for thick shields (9). It is clear that for this flare a dose of 100 rads would be received under a composite shield of 6 gm/cm^2 of aluminum and 25 gm/cm^2 of polyethylene. The magnitude of this dose value indicates that for a flare with the characteristic of the 14 July 1959 event, considerable shielding will be required on lunar missions.

SUMMARY

The computer program, STREP, may be utilized as a generation tool for a source of time-integrated spectra incident on a space vehicle during cislunar missions. Since input requirements on the trajectory are positions and associated times, the procedure is independent of a particular calculational method for trajectories. The results from problems which have been considered to date indicate that the present version is adequate for lunar missions, provided a radiation field does not exist in the vicinity of the moon. Currently, an adaptation of the procedure is under development to handle interplanetary missions.

The calculations for a lunar mission during the active period of a particular solar flare indicate an excessively high radiation dose will be received within an unshielded vehicle. The addition of sufficient shielding to lower the dose to 100 rads creates a hazard from secondary radiations born in the shield. These results point out the necessity of planning missions during quiet solar periods to avoid excessive shield weights.

REFERENCES

1. Van Allen, J. A., "The Geomagnetically Trapped Corpuscular Radiation," Science in Space, Berkner and Odishaw, Editors, New York, 1961, p. 275.

2. Freden, S. C., and R. S. White, "Particle Fluxes in the Inner Radiation Belt," J. Geophys. Res., vol. 65, 1960, p. 1377.

3. Holly, F. E., L. Allen, and R. G. Johnson, "Radiation Measurements to 1500 km Altitude at Equatorial Latitudes," J. Geophys. Res., vol. 66, 1961, p. 1627.

4. Winckler, J. R., P. D. Bhavsar, and L. Peterson, "The Time Variations of Solar Cosmic Rays During July 1959 at Minneapolis," J. Geophys. Res., vol. 66, 1961, p. 995.

5. Biswas, S., C. E. Fichtel, and D. G. Guss, "Solar Protons, Alpha Particles, and Heavy Nuclei from November 12, 1960 Flare," AGU meeting, UCLA, 27-29 December 1961.

6. McCracken, K. G., "The Cosmic Ray Flare Effect, The Flare Effects of May 4, November 12, and November 15, 1960." J. Geophys. Res., vol. 67, 1962, p. 435.

7. Chadwick, F. W., and C. E. Herrick, "The Two-Body Required Velocity Program," Convair/Astronautics ERR-AN-022, 7 October 1960. (U)

8. Gapcynski, J. P., and D. S. Woolston, "Characteristics of Three Precision Circumlunar Trajectories for the Year 1968," NASA-TN-D-1028, March 1962. (U)

9. Wilson, R. K., R. A. Miller, and R. L. Kloster, "A Study of Space Radiation Shielding Problems for Manned Vehicles," General Dynamics/Fort Worth FZK-144, 8 June 1962. (U)

RECOVERABLE SPACE LAUNCHING NUCLEAR SYSTEMS

Maxwell W. Hunter, Jr.
National Aeronautics and Space Council
Washington, D. C.

ABSTRACT

Factors affecting the use of nuclear rockets in recoverable space launching systems are discussed. Energy requirements and consequently fuel costs are shown to be of the same magnitude as for terrestrial air transport systems. The ability of advanced nuclear propulsion to permit the recovery and re-use of airframes, and aid in the achievement of low refurbishment costs is discussed. Attention is directed toward certain items of safety, including some basic factors which might cause gaseous fission rockets to be more operationally safe than solid core units. The possible impact of economical space launching systems on national programs is considered.

RECOVERABLE SPACE LAUNCHING NUCLEAR SYSTEMS

Maxwell W. Hunter, Jr.
National Aeronautics and Space Council
Washington, D. C.

INTRODUCTION

To increase the utility and decrease the cost of space launching systems would constitute major progress in our national space program. The basic approaches toward this objective are to reduce the cost of the fuel used in current systems, either by reducing its unit cost, its quantity, or both, to reduce the price of the hardware required, and to reduce the cost of launch operations and facilities. Nuclear rockets offer a means of reducing fuel costs and hardware cost reduction almost certainly requires the recovery and re-use of the equipment. Hence, questions of reliability of recovery and cost of refurbishment may become the key items. Launch operation improvements require a new class of more convenient vehicles.

This paper is restricted to examining the effect of nuclear propulsion upon the cost factors involved in the delivery of payloads to low earth orbital altitudes, although the required velocity is low in terms of the actual potential of nuclear propulsion. The utilization of both early availability low performance nuclear systems, and more optimum advanced systems will be considered.

ENERGY REQUIREMENTS AND FUEL COSTS

The total impulsive velocity required to place objects in low earth orbit, with appropriate allowances for drag and gravity losses, is of the order of 30,000 fps. It is not the intent of this paper to present detailed trajectory calculations to define accurately this number, but rather to delineate the basic characteristics of vehicles operating approximately in that region. Since this paper will be dealing with very high energy propulsion systems, however, a total impulsive velocity of 60,000 fps, enough to permit re-entry and landing by rocket retro thrust only, will also be considered.

The propellant to launch gross weight ratios required for single-stage vehicles for the two values of impulsive velocity mentioned, are shown as a

function of specific impulse in Figure 1. A value of 900 seconds specific impulse is typical of an advanced solid core nuclear thermal rocket; values much higher than that will probably require some form of gaseous fission rocket. (Liquid core fission rockets are a possibility, but show only modest improvement over solid core systems compared to the latent potential of gaseous fission systems.) If gaseous fission systems can be made to work at all, they should yield specific impulse values on the order of 2500 seconds using only regenerative cooling. They will in this region have thrust-weight ratios high enough to permit take-off from the Earth. For higher performance, radiators are required to reject excessive heat. The deterioration of thrust-weight ratio in this region is subject to considerable question. (Reference 1).

It can be seen from Figure 1 that the amount of fuel which must be used, even when re-entry is done completely by rocket braking, is a remarkably low value at high specific impulses. The fuel load actually can be lower than that currently carried by transcontinental jet transport aircraft. Although this seems startling, the energy requirements are not really much different. If one assumes a jet transport to be cruising for 6 hours at an L/D of 12, this means that the engines were applying 1/12 of the weight of the airplane for the six hour period. If this could have been applied in field-free space, the vehicle would have accelerated 1/12 of a g for 6 hours, and would have generated 58,000 fps, essentially the maximum total velocity considered here. Thus normal cruising aircraft utilize the same order of magnitude of energy as required for a modest space transport system, essentially because such aircraft fight gravity incessantly during their whole flight. In space operations, high thrust rockets fight gravity quickly and efficiently in the early part of their flight, and coast afterwards. Space flight is expensive, hence, not because the energy required is high, but because no way has yet been found to package and utilize these energies out in space. Only the nuclear rocket achieves high enough exhaust velocity to create the necessary performance in a single-stage vehicle, and can even reduce fuel load requirements to the same magnitude as those of conventional transport aircraft.

Although Figure 1 indicates that the highest specific impulse obtainable is the most desirable, this is not true from the cost viewpoint. High specific impulse requires the burning of more nuclear fuel per unit of propellant expelled in the exhaust and, therefore, the total unit cost of propellant increases. Figure 2 shows total propellant cost as a function of specific impulse for uranium fuel and hydrogen propellant. Even up to a specific impulse of 10,000 seconds, the effect of increased uranium consumption is modest and does not substantially affect design considerations. A factor which does strongly effect the propellant cost for high specific impulse is that gaseous fission systems will probably not have perfect

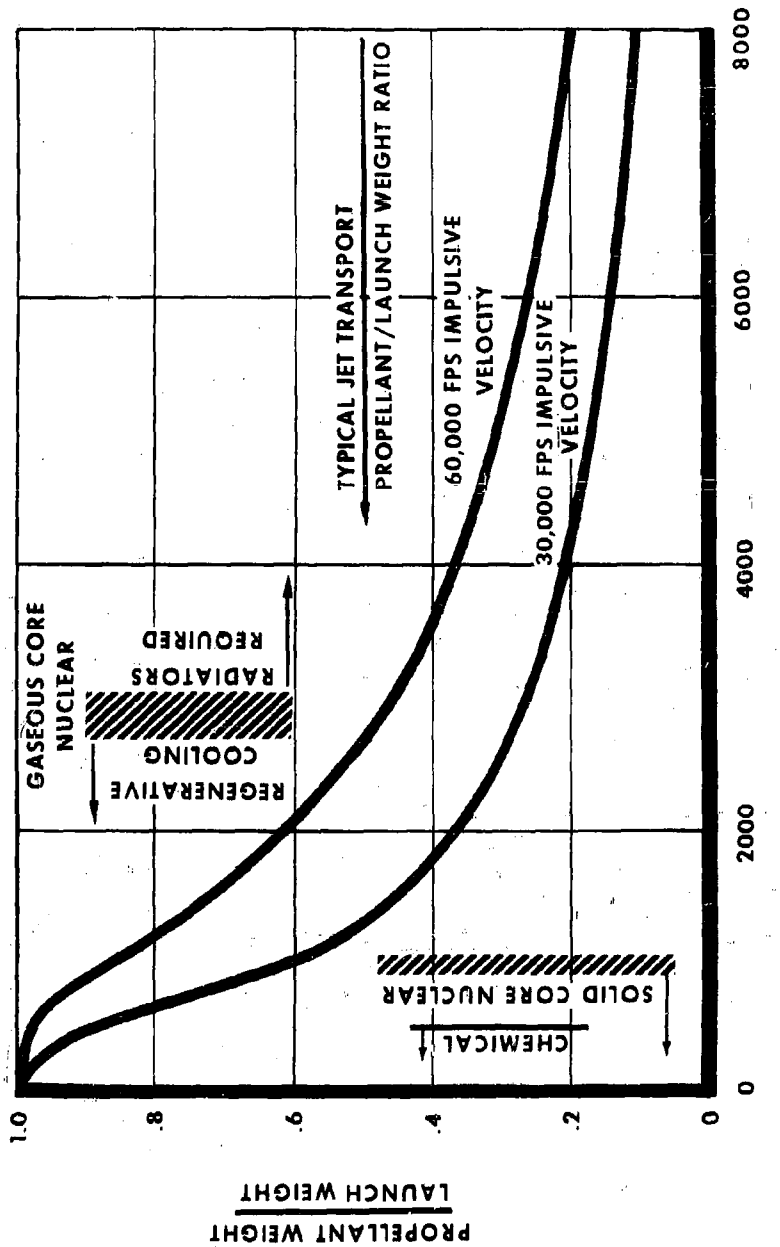


Fig. 1. Required Fuel Weights for Space Launch Vehicles, Single Stage

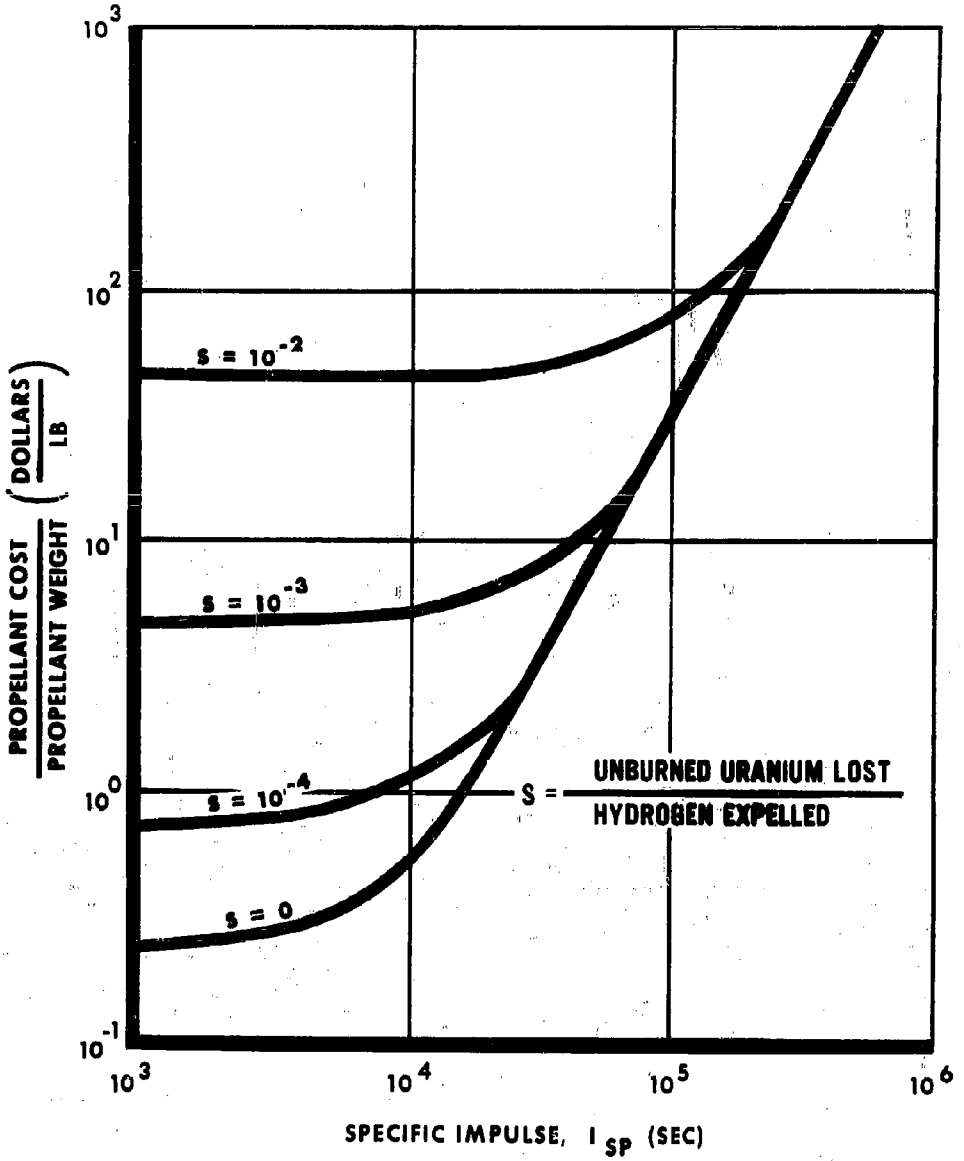


Fig. 2. Cost of Propellant for Fission Reactor Rocket

fuel containment. Therefore, the increase in propellant cost for several separation ratios is also shown in Figure 2, and a curve of propellant cost as a function of separation ratio is shown in Figure 3 (for Isp=3000 seconds). It is believed that a separation ratio of 10^{-3} is certainly attainable and 10^{-4} possible. The propellant cost per unit launch weight is shown in Figure 4. This cost is quite reasonable and, furthermore, the cost of providing complete propulsive braking is not large once suitable engines become available. Hence, the key to achieving low cost operations is the design use of the large weight fractions available to achieve the best trade-off between payload delivered to orbit and structural weights devoted to enhancing long life, re-use, and convenience of operation.

AIRFRAME COSTS

Experience has shown clearly that structures which fly through the air or into space are expensive. At times, structures such as throw-away tip tanks or booster cases have been made for a reasonably low price, and the utilization of relatively inexpensive throw-away structures as a cost reduction method consequently cannot be ignored. The most commonly utilized cost reduction technique however, is to re-use the vehicles. There are two basic reasons for this. The first is that only a small fraction of the expense of an airframe can be made relatively inexpensive. There are always many items of inherently expensive equipment on board, including engines, guidance, life support, etc., which must be re-used to lower costs. Since the re-use technique must be applied anyway to most of the equipment, it is normal aircraft practice to simply apply it to everything. The second reason is that it has been possible in aircraft practice to attain both essentially perfect recovery reliability, and also almost zero refurbishment costs between flights. Thus, otherwise uneconomical costs have been successfully diminished in aircraft practice.

The effect on cost of number of re-uses with zero refurbishment and perfect recovery reliability is shown in Figure 5, which is plotted from data of Reference 2. These curves indicate the expected drastic decrease in cost by utilization of such techniques. The usual space vehicle launch analyses do not yield costs nearly this low for several reasons. In the first place, chemical boosters are normally considered and their size is such that costs are high even under these assumptions. Secondly, it is assumed that the structures required to meet the severe environmental factors would inevitably lead to expensive refurbishment costs. In addition, since the performance of the basic vehicle usually does not enable it to go into orbit and land later at its original launching base, it is assumed that the vehicle will land at some remote point, have to be returned to the launching base, and quite possibly be reconditioned after adverse treatment during landing.

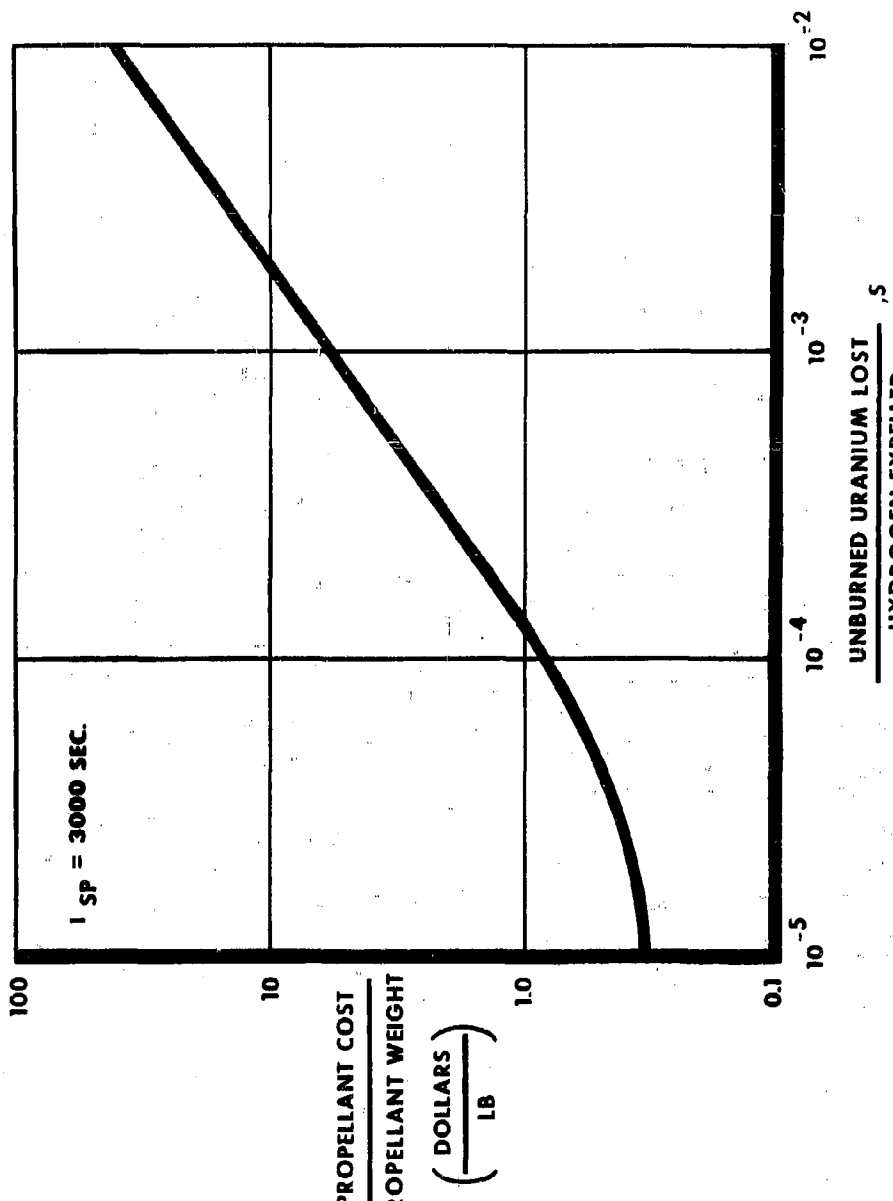


Fig. 3. Cost of Propellant for Gaseous Fission Rocket

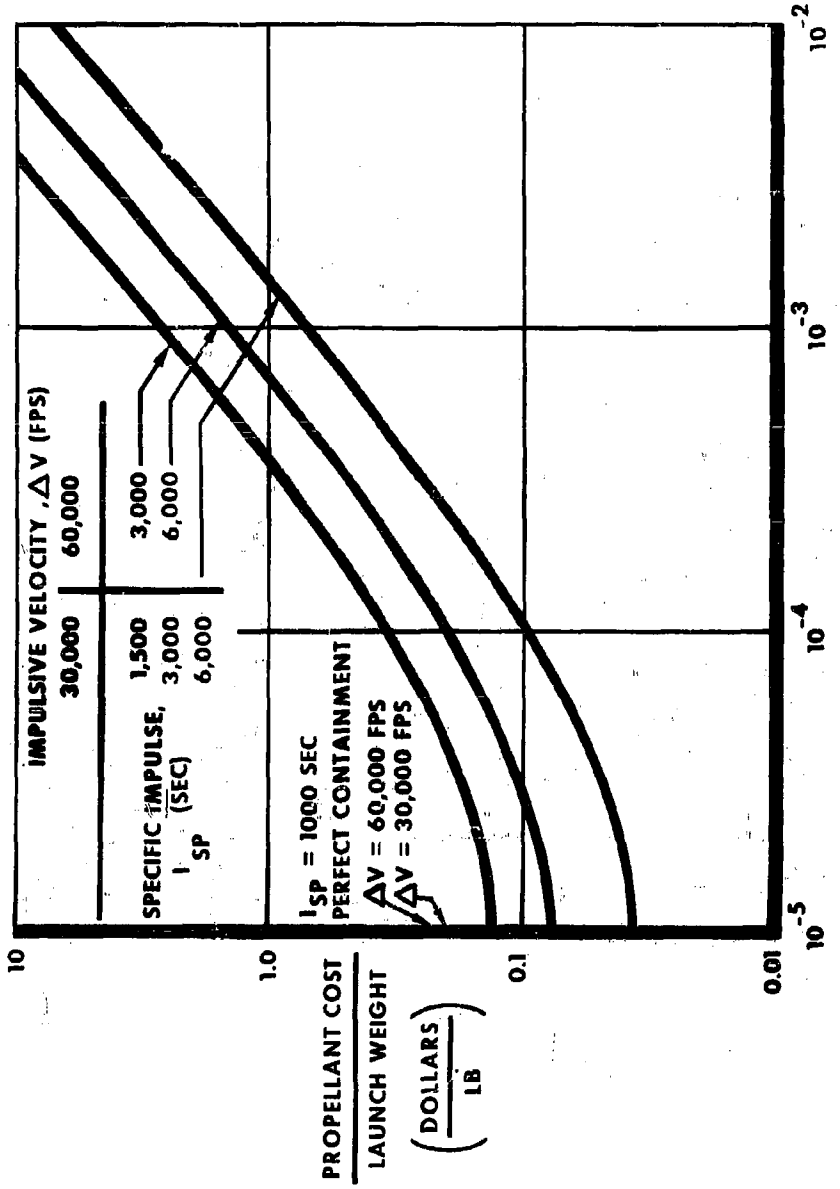


Fig. 4. Propellant Cost per Unit Launch Weight

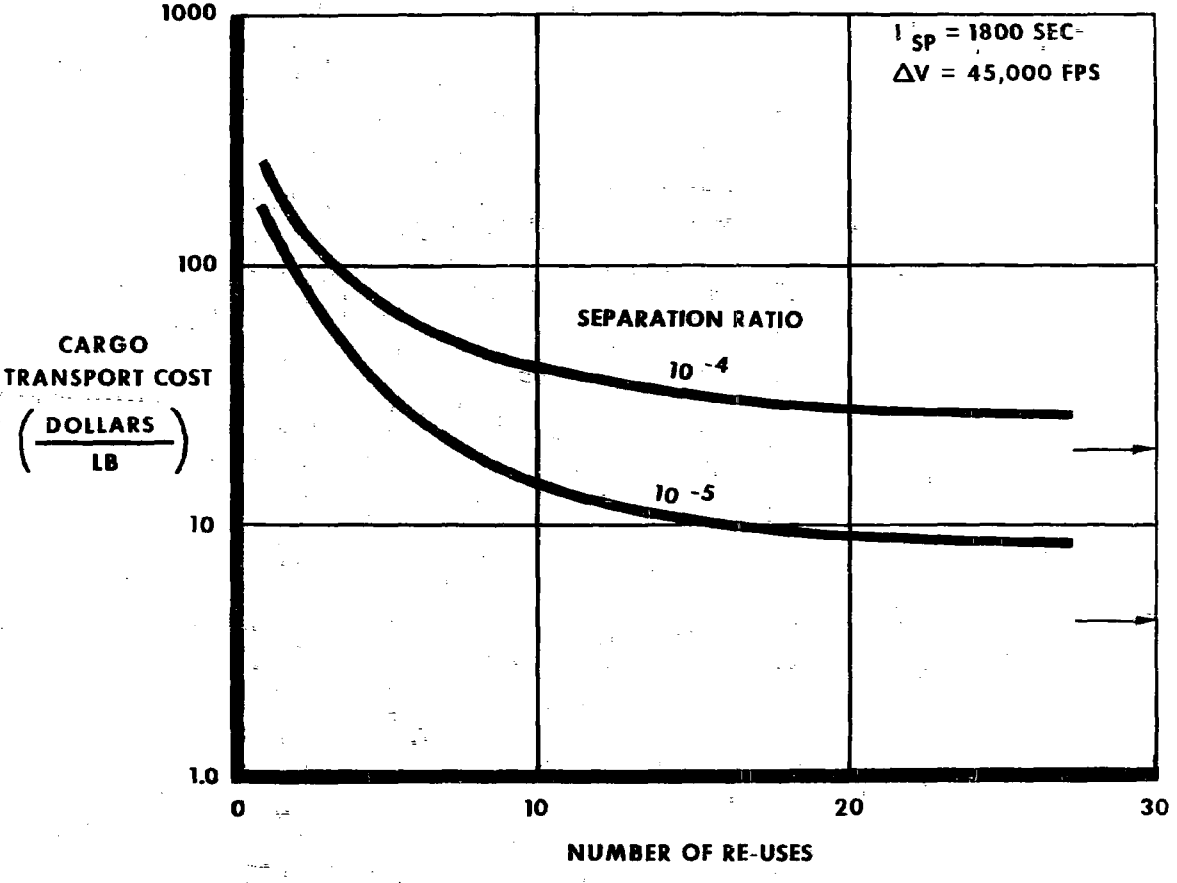


Fig. 5. Effect of Reuse on Total Cost, Zero Refurbishment Cost

Most space launch vehicle analyses are based on the experience of the missiles industry. The total recovery and re-use experience of most missiles engineers can be summarized by stating that it is a rare and pleasing experience whenever any hardware is even available to look at afterwards. This, it is submitted, is not conducive to a good intuitive grasp of the basic factors in recovery reliability, let alone re-use. In general, when discussing maintenance of re-usable space vehicles, the point is made that large refurbishment costs will be required between flights due to the extreme environmental problems faced by re-entering vehicles. One certain solution, in case technology is not able to meet the challenge of direct assault on the environmental problems, is simply to provide excessive propulsive capabilities which are used to effect re-entry under operational environments no worse than those familiar in normal aircraft practice. It is hoped that such a drastic solution will not really be necessary, but nuclear propulsion, in addition to its other advantages, will make this solution available if aerodynamic and structural design practice fails. It is the duty of system engineers to realize the potential of this development and press for its implementation. There has already been too much time and energy wasted over the alleged fantastic fundamental expense of space flight.

The point is frequently made that massive space operations are necessary to justify the development of recoverable systems. The rationale of this is somewhat evasive. It is true that a certain amount of extra work is required to make a vehicle recoverable. It is equally true, however, that even during a development program, the production of a smaller number of vehicles tends to cut costs. The ability to have vehicles available for re-examination after flight greatly accelerates reliable equipment development, and the confidence inspired in the crews by the ability to abort flights safely in case of early difficulty, should greatly accelerate vehicle development and acceptance. The often-stated argument that a non-recoverable vehicle can be developed cheaper than a recoverable one has never been convincing.

SAFETY

Nuclear rockets present a substantial safety problem in use. Actually, almost any high energy device requires substantial safety procedures. If one had to choose between being hit on the head with a jet airplane loaded with several hundred thousand pounds of kerosene, or a nuclear rocket loaded with several hundred thousand pounds of hydrogen, it is doubtful that a very great preference would ensue. The additional concern with nuclear rockets, of course, is the extra technological and psychological problems surrounding any use of nuclear energy. While no attempt will be made to examine fully the safety aspects of nuclear rockets, several

points will be mentioned.

It is by now apparent that nuclear rocket engines are not bombs, that the possibilities of critical excursions are very low, and that the resulting release of fission products would be equivalent to that of only a fractional yield bomb. Thus, even in the event of an accident on the launching pad, the resulting debris can be characterized more as an awkward clean-up problem than as a major catastrophe. The radiation flux surrounding even a large nuclear rocket on direct take-off from the ground is such that a person unprotected only one mile away would receive less than one rem dosage. Compare this one mile radiation exclusion radius with the fifteen mile noise exclusion radius of a chemical NOVA class vehicle. Thus, nuclear vehicles could take off from relatively normal launch bases and airfields, and proceed throughout most of their burning phase under normal launch safety control procedures. There is a time when the vehicle's velocity exceeds ICBM speed, but is still sub-orbital so that a vehicle flight abort might cause impact anywhere on the planet. This particular critical failure time can be handled by auxiliary chemical propulsion capabilities used either to decrease the velocity so that the vehicle will fall within the range, or increase it to an orbital speed. If the trajectory is held unusually low - a desirable thing anyway from flight abort considerations - the critical time is only a few seconds and only a few hundred fps retro velocity must be supplied by the emergency chemical system.

Once the vehicle is well up into the atmosphere, negligible danger would exist if it were possible to deposit the fission products in the atmosphere. Hence, provision for burn-up of the fuel elements on atmospheric re-entry would present a basic solution. It is quite conceivable that this would work well, and auxiliary power units are now being designed this way. Considerable effort might be required to insure positive burn-up of a large solid core nuclear rocket, however. In the case of gaseous fission rockets, all active nuclear fuel is in the gaseous phase and is easily and safely injected into the atmosphere in vapor form in case of emergencies. In fact, it is probably impossible to prevent its injection in case of trouble.

Intuitive reasoning would say that the gaseous fission rocket is more dangerous than the solid core fission rocket since its containment would not be perfect. The opposite is probably true. As indicated above, it may be difficult to insure emergency burn-up with solid fuel elements. Furthermore, the fission product build-up continues in solid cores as number of re-uses increases, and this inventory must be returned to low altitudes after each flight. Thus atmospheric disposal is complicated, and the vehicles require special handling upon return. Gaseous systems would be expected to dispose of their fission products in the high atmosphere

after each flight so that the returning vehicles are as safe and as easy to approach and service as chemical vehicles. It therefore follows that each succeeding take-off would start with zero fission products aboard. The take-off could be made perfectly safe if auxiliary chemical propulsion were utilized. Estimating the amount of such auxiliary propulsion requires a reasonably sophisticated analysis, and the following discussion is based on Reference 3. If we strive for a very high degree of safety, then an accident which causes the release of all fission products aboard must be analyzed by the usual atmospheric dispersion techniques, and the dose to an exposed person calculated at the worst location on the Earth's surface. In general, the worst flight time for such an accident will not be immediately at reactor ignition since no fission product build-up has occurred. Nor will it be at end of burning, since although fission product inventory is presumably highest, atmospheric dispersal is very effective. Hence, one must assume different start altitudes and then make calculations for accidents at various burning times until the worst case is located. The results of such a set of calculations (from Reference 3) are shown in Figure 6. It is noteworthy that even for six million pound thrust rockets, only about four thousand feet start altitude is required for complete safety.

The low safe altitudes indicated are startling. They show simply that atmospheric dispersion is very effective even at low altitudes, and that only small amounts of fission products are ever aboard to be dispersed. Hence, not only must one remember that propulsion reactors are not bombs, but also that gaseous fission reactors are not as dangerous as normal power or propulsion reactors since, in space operations, they will not build up fission products aboard ship. The philosophical and technical objection to flying reactors, which is basically because of the fission product load aboard, is thus almost completely discounted. The few thousand feet safe altitude can be obtained with the order of a few hundred fps of chemical boost, with a resultant total vehicle launch weight increase of less than 10%. It might also conceivably be obtained by airplane-like techniques, since the low fuel loads required for orbital transports indicated in Figure 1 allow airplane-like weight penalties.

IMPORTANCE TO NATION

It should be clear that the achievement of substantial cost reductions in space transportation systems can be of extreme importance to the national space operations, both scientific and military. The intellectually easy way of doing space flight, i. e., with large throw-away chemical rockets such as are being developed today, tends to yield costs of delivery of cargo to orbit that are three orders of magnitude higher than the delivery of such cargo to points on this planet's surface. Any new space missions to be undertaken, whether lunar and planetary explorations, whether

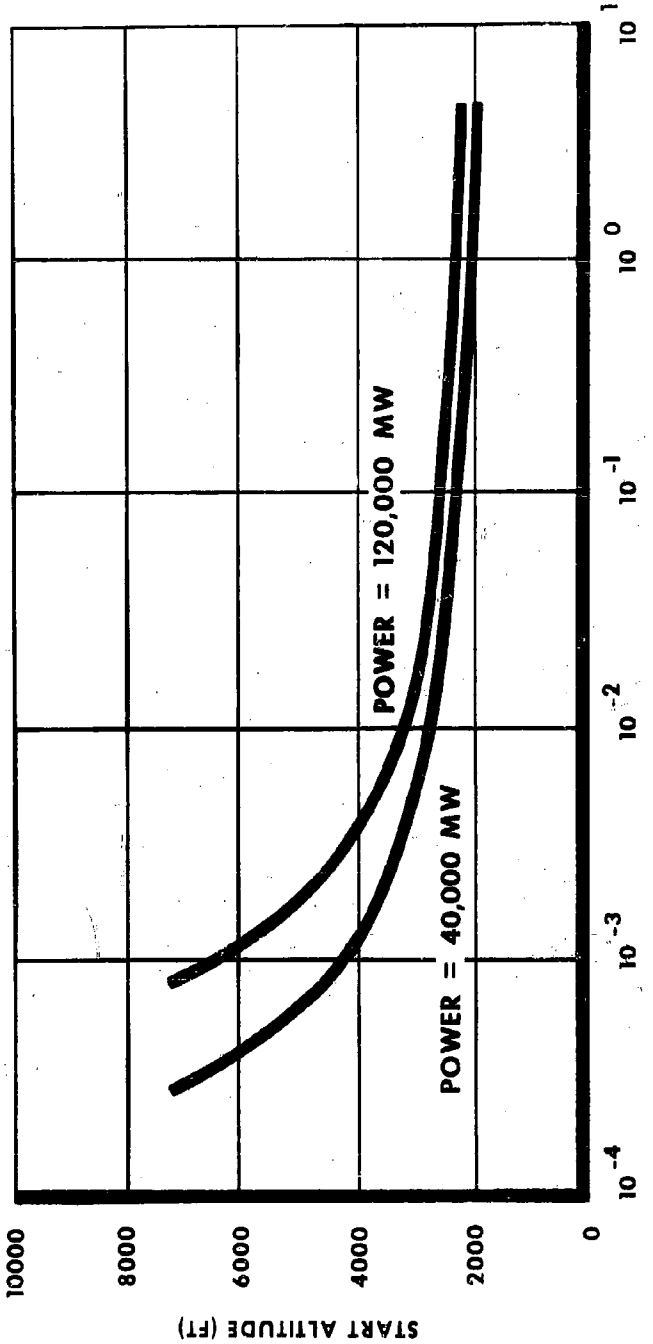


Fig. 6. Dose to Ground Observer vs. Start Altitude for Gaseous Core, Dosage Selected for Most Critical Time of Gaseous Core Operation

military or non-military operations, are viewed in the context of this fantastic technical cost barrier. This factor can be crippling.

It is commonly supposed that it is possible to be logical in the analysis of new missions. That, for instance, a new military system can be subjected to an advanced operational analysis which determines logically the desirability of its development. It is the opinion of the author that this is rarely true. Too many intangibles exist which are not subject to operational analyses. Even in this modern day, experienced intuition is a strong factor (which is freely wielded). Intuition and, more importantly, imaginative visualization of new possibilities which are not obvious in logical engineering analysis, are greatly aided when people with vision and imagination are placed where these gifts can be exercised. It is doubtful that any group of people, however intelligent, sitting around a conference table on planet Earth, can adequately grasp the military capability of bases on other planets, particularly when they have never even been close to the Moon. Nor can they envision the operational strategy and tactics of fleets of vehicles operating in deep space environments where detection and kill mechanisms are so foreign to all previous experience. Geocentric arrogance afflicts our modern technology as it once did, and sometimes still does, religion.

The question of the desirability of development of new systems is greatly modified if the basic cost permits relatively economic experimentation, rather than if total operational effectiveness must be assured to justify any experiment at all. Thus, a three order of magnitude reduction in space transportation costs would likely become the key item in our whole decision process with respect to future space programs of all types. If this reduction is at all feasible, it will be an inexcusable engineering failure if we do not pursue whatever techniques are necessary to materialize this technical possibility. Apparently, as in almost all transportation problems, advanced propulsion is the key to real improvement. Total understanding of intermediate propulsion techniques might be a waste of time.

CONCLUSIONS

Nuclear propulsion offers a key to vastly more economical space launch systems.

The energy requirements for orbital flight are no larger than for current Earth air transport systems.

Re-use of vehicles is of key importance. Advanced nuclear rockets reduce fuel load requirements so that single-stage vehicles can be used. Inefficient structures and even rocket retro braking to avoid difficult

re-entry environments may be used, if necessary to achieve small refurbishment costs.

Nuclear rockets can be made safe in operation. Gaseous fission rockets, in particular, can avoid fission product build-up by disposal in space. Hence, the objection to over-flight with fission products is avoided and convenient handling after flight becomes possible.

Major cost reductions in space launch systems would have a very strong effect on national space decisions.

REFERENCES

1. Hunter, M. W., Jr., "The Potential for Nuclear Propulsion for Manned Space Flights," presented at the Institute of the Aerospace Sciences, St. Louis, Missouri, April 30-May 20, 1962.

2. Hallet, R. W., Jr., McKee, J. W., Faller, R. J., "Operational Use of Gas Core Nuclear Reactor Propulsion Systems in Large, Manned, High-thrust Spacecraft." Douglas Aircraft Engineering Paper No. 1200. Presented at 10th Annual Meeting of the German Rocket Society, Bremen, West Germany, September 23-25, 1961.

3. Hallet, R. W., Jr., "RITA: 'Scheduled' Mars and Venus Trips by Single-stage Nuclear Vehicle," prepared for publication in Space/Aeronautics R&D Technical Handbook (1962-1963).

REUSEABILITY AND DISPOSAL OF NUCLEAR SYSTEMS IN SPACE

A. L. Bethel
and
J. H. Bach

Westinghouse Electric Corporation
Astronuclear Laboratory
Large, Pennsylvania

ABSTRACT

The problems of maintaining, refueling, and disposal of radioactive nuclear space systems are considered. The types and frequencies of different maintenance operations are estimated, as well as special problems affecting military systems. Design requirements to permit maintenance are outlined for propulsion and APU reactors, together with a number of service schemes. Secondary applications for used radioactive components are described and specific disposal orbits are recommended based on fission product activity, and Van Allen Belt location.

The work described in this paper was company sponsored.

REUSEABILITY AND DISPOSAL OF NUCLEAR SYSTEMS IN SPACE

A. L. Bethel and J. H. Bach
Westinghouse Electric Corporation
Astronuclear Laboratory
Large, Pennsylvania

INTRODUCTION

Operational experience with nuclear power systems on earth shows that they are reliable, dependable, easy to operate and, in general, more satisfactory in many respects than the designers anticipated. We expect the same general pattern to obtain in nuclear space systems and therefore believe that there will be in the next 10 to 15 years a great many nuclear propulsion and power systems at or near the point of operational use.

On the other hand, as no surprise, refueling, maintenance, and repair of nuclear plants has taught us that they are somewhat more complicated to repair and maintain, that refueling systems leave much room for improvement, and that spent fuel and other radioactive debris from the plants are a considerable nuisance. It is timely therefore, to begin solving these problems now, along with development of nuclear space systems so that in a real sense, adequate refueling, maintenance and repair capabilities can be built into each system.

This paper is an attempt to stimulate thinking on the following questions as they relate to nuclear space systems:

1. What items will require replacement and what is their replacement frequency?

2. How can this be done in view of the hazards involved and what additional equipment is required?

3. What is the value, and how does one make use of "secondhand" radioactive components?

4. Are nuclear systems particularly vulnerable to space — military environments?

While any military space system must eventually face the maintenance-in-space problem, nuclear systems pose special difficulties. Nuclear fuels, while only

slightly radioactive before use, are highly radioactive subsequent to reactor operation. Nuclear fuels, new or used, cannot be stored or handled indiscriminately due to their inherent reactivity, and care must be taken to avoid accidental criticality. Components which do not contain nuclear fuel but are located in the vicinity of a reactor are activated by the neutrons emanating from the core during reactor operation and remain radioactive for some time thereafter depending on the decay constant of the specific isotopes.

The extreme susceptibility of biological systems to ionizing radiation demands effective shielding and hence limits maintenance to remote systems. Even inert materials, particularly electronic components must be protected if they are to operate reliably over any extended period of time.

On the other hand, undamaged, partially depleted fuel elements and other radioactive materials are potential assets because they are an energy source. Components removed from unattended systems in the interest of reliability still have functional value and might, for example, be useful in a situation where day-to-day maintenance is practicable.

What is needed, obviously, is a means of utilizing the attributes of the material at little additional cost, and of placing the unuseable material in some convenient location where it presents no hazard either to earth or to those in space. We do not have a simple, elegant solution to these problems in this paper.

We have addressed ourselves to a limited number of specific problems. Typical and very general cases have been selected for analysis and where solutions are suggested, these have been treated in a very brief manner indeed. We shall start by stating some basic assumptions and then discuss briefly the following items:

1. Maintenance Requirements and Frequency
2. Service Facilities
3. Nuclear System Design Requirements
4. Secondary Application
5. Final Disposal.

ASSUMPTIONS

In order to discuss these problems at all, one must make some basic assumptions with respect to the nature and employment of nuclear plants in space. For military purposes we stipulate three types of nuclear systems which will be used:

1. Nuclear propulsion plants for space vehicles

2. Auxiliary power units to provide service power for space craft and satellites

3. Power plants for permanent orbiting, lunar, or planetary stations.

We further envision that most space craft, once in space, will remain there permanently and will be re-manned and re-supplied by logistic space vehicles. Further, we shall assume that lunar or planet based vehicles will eventually draw upon local raw materials for propellants. This does not assume necessarily that H_2 will be found on the moon, but it puts some additional burden on future engine designs.

MAINTENANCE REQUIREMENTS AND FREQUENCY

Nuclear systems for military applications will require maintenance for two reasons. First, routine service and maintenance for normal wear including propellant depletion, parts failures, nozzle wear, and core burn up. The second major cause for repair will be due to damage by the space environment or enemy action.

Routine Maintenance

Requirements for routine replacements are difficult to establish until some operating experience has been accumulated; but by estimating core refueling frequencies maximum permissible replacement schedules for other components can be established.

The frequency with which propulsion cores must be replaced due to fuel depletion and fission product poisoning is a function of both neutron physics and operational considerations. Consider a thermal reactor rated at 2,000 MW(th) with a 200 kg U^{235} loading, and let us assume that after 4% U^{235} burnup the core can no longer override equilibrium Xenon and thus loses some of its military utility. If one examines a wide range of mission requirements which include orbit transfers, acceleration to escape velocity, deorbiting, rendezvous, and landing maneuvers, the overall propulsion operational periods will average 15 minutes each, and one such period every 72 hours constitutes an average demand cycle. Such a core would then have to be replaced every 3 years and would have actually operated 96 hours. This latter number points up the requirement that nozzles should last on the order of 30 hours so that the propulsion plant need not be touched at all more than once a year; there should be no question about pumps and electronic systems capable of such use periods despite the 400 startups and shutdowns presumed by this duty cycle.

Besides reactivity changes due to burnup, more subtle processes will also affect reactor reactivity. The propellant might diffuse into the core moderator or, in case of fast reactors, into the matrix forming compounds which cause corrosion or which are blown out through the nozzle at subsequent startups. Depending upon the nature of these reactions, they will increase or decrease reactivity and in all likelihood a number of these will occur simultaneously tending to offset each other to some degree.

One may now propose acceptable replacement schedules for other nuclear system components and at the same time one can estimate the cost of transporting these into space.

Table 1 gives major radioactive components, their unit weights, followed by the estimated full power life expectancy of the item in column 3, and in column 4 the required total operating time for the specific component over the vehicle life. We shall assume a strategically useful vehicle life of ten years. Koelle (1) recently estimated future space payload costs, and we shall assume a permissible military transportation cost five times his maximum acceptable commercial cost for lunar missions; transportation for replacement components will therefore cost $125/lb of useful payload. It can be seen that the nuclear component space transportation cost over the life of the vehicle is about $600,000/year. This does not seem excessive compared with the probable first cost of the vehicle or the expected manufacturing costs of the components themselves.

Environmental And Enemy Damage

Apart from normal equipment malfunction problems, military space reactors may be exposed to three hostile forces: meteorites, proton showers and enemy action.

Meteorite damage danger does not seem any greater for nuclear than non-nuclear systems. A meteorite puncturing a fuel tank, or damaging a heat exchanger would presumably cause the same amount of inconvenience, and hence the same amount of protection and repair capability is required. One special facet should be mentioned: loss of primary coolant from a closed reactor system due to meteorite puncture may cause the core to burn out unless both automatic isolation valves and a separate decay cooling system are provided. As will be shown later, the cooling system should be available anyhow to permit normal repair and maintenance of primary system components.

Solar flares are probably of greater significance to military than to commercial missions since the former cannot always be timed to coincide with favorable astronomical periods. This factor removes some of the onus of the shielding penalty usually assigned to nuclear systems when compared to other power sources. While biological shielding against primary protons seems prohibitive, (except possibly by magnetic deflection), the secondaries can cause as much or even more damage, so that the crew must be shielded whether or not nuclear systems are employed. Proton showers do not appear to be of sufficient intensity to cause significant interaction with the reactor neutronics and should not precipitate reactor control problems. They might, however, cause trouble in the following manner: MHD engines will require magnetic fields which, for space vehicles, can only be provided by superconducting magnets. If a proton or its secondary happens to strike a superconducting wire causing a local temperature rise, at the same time generating other secondaries which will cause additional temperature increases in the same locality, the resistance in the wire will rise very quickly and the field will be destroyed. We are presently investigating whether these fields are sufficiently strong to serve as proton deflectors.

Table 1. Replacement Economics for Nuclear Space Components

Item	Wt (lbs)	Full Power Life (hrs)	Operation over Vehicle Life* (hrs)	Replacements over Vehicle Life (No)	Unit Space Cost*** ($ X 10^{-3})
2,000 MW(th) Propulsion Core	1,100	100	300	3	137
100 e KW APU Reactor	400	15,000	57,000**	4	50
Pump (Propulsion)	2,500	500	900	2	312
Pump (APU)	50	25,000	88,000	4	6
Nozzle	3,000	30	300	10	370
Nuclear Control (Propulsion)	4,000	200	300	1.5	500

Total Space Replacement Cost (excl. of mfg. cost): $5.7 Million

*Strategic Useful Vehicle Life 10 Years

**65% Load Factor

***$125/Payload lb Space Transport Cost Assumed

The ability of nuclear space systems to withstand enemy action requires special examination in at least one area. Nuclear explosions in space will not cause a blast effect but depend on heat and radiation for their damage. Neutron showers from such explosions are not attenuated except by the inverse square effect. Hence, a 0.5 megaton burst 30 miles away will cause local neutron flux peaking of appreciable magnitude in a typical propulsion reactor. Fast reactors would appear to be more vulnerable, since a step increase in fast neutrons would cause the delayed neutron fraction to vanish. We do not contend that such an event would be fatal, but rather that the possible effects of nuclear explosions upon specific space reactors require detailed examination.

SERVICE FACILITY REQUIREMENTS

Once one knows something about the nature of repairs, the necessary facilities can be postulated. Military use of nuclear space systems will require facilities capable of handling highly radioactive components. Service concepts for earth-based vehicles, space maintenance ships, and on-board repairs will be discussed here briefly.

Earth-Based Vehicles

Handling facilities for those vehicles which must per force shuttle between earth and space might be built as shown in Figure 1. Here we assume a nuclear aerospace plane which takes off and lands in the atmosphere using aerodynamic lift and chemical engines; the nuclear reactor is brought critical only at high altitudes. Upon landing, the reactor which is located on an extendable structure, is lowered into a shielding facility such as a pool, and the craft can now be serviced with relatively small inconvenience as far as shielding requirements go. The reactor and nozzle can be detached under water and serviced in conventional hot cells or be replaced with a new unit.

Of course, this does not solve the one problem which has to date kept nuclear powered aerospace planes limited to drawing board concepts: How does one handle a take off or landing accident which might contaminate the whole air base and render it useless for further operations? It seems quite clear that few, if any of present bases can be used for such ships; at the minimum, sites similar to Edwards AFB will be required. Even here any approach pattern must pass over populated areas; and when one incorporates the possibility of an air collision into the maximum credible accident analysis, almost all inland bases are ruled out. One solution would be to locate the bases on the edge of large bodies of water so that any abort is likely to end up in the water, thus providing dilution for dispersed fission products. Since islands must always be approached by water, small islands located in close proximity to each other, to provide alternate landing strips in the event of contamination, might prove of strategic importance. An alternate approach would be the use of seaplanes. An aerospace plane based on this concept was proposed some time ago by the authors but was considered impractical by aircraft designers due to the high structural stresses imposed during take off and landing. However, modern hydrofoil or hovercraft designs might overcome this problem. Water basing would also simplify the protective measures which must be taken to prevent gamma irradiation incidents in the area swept by the solid angle of a decaying reactor in the approach or take-off pattern.

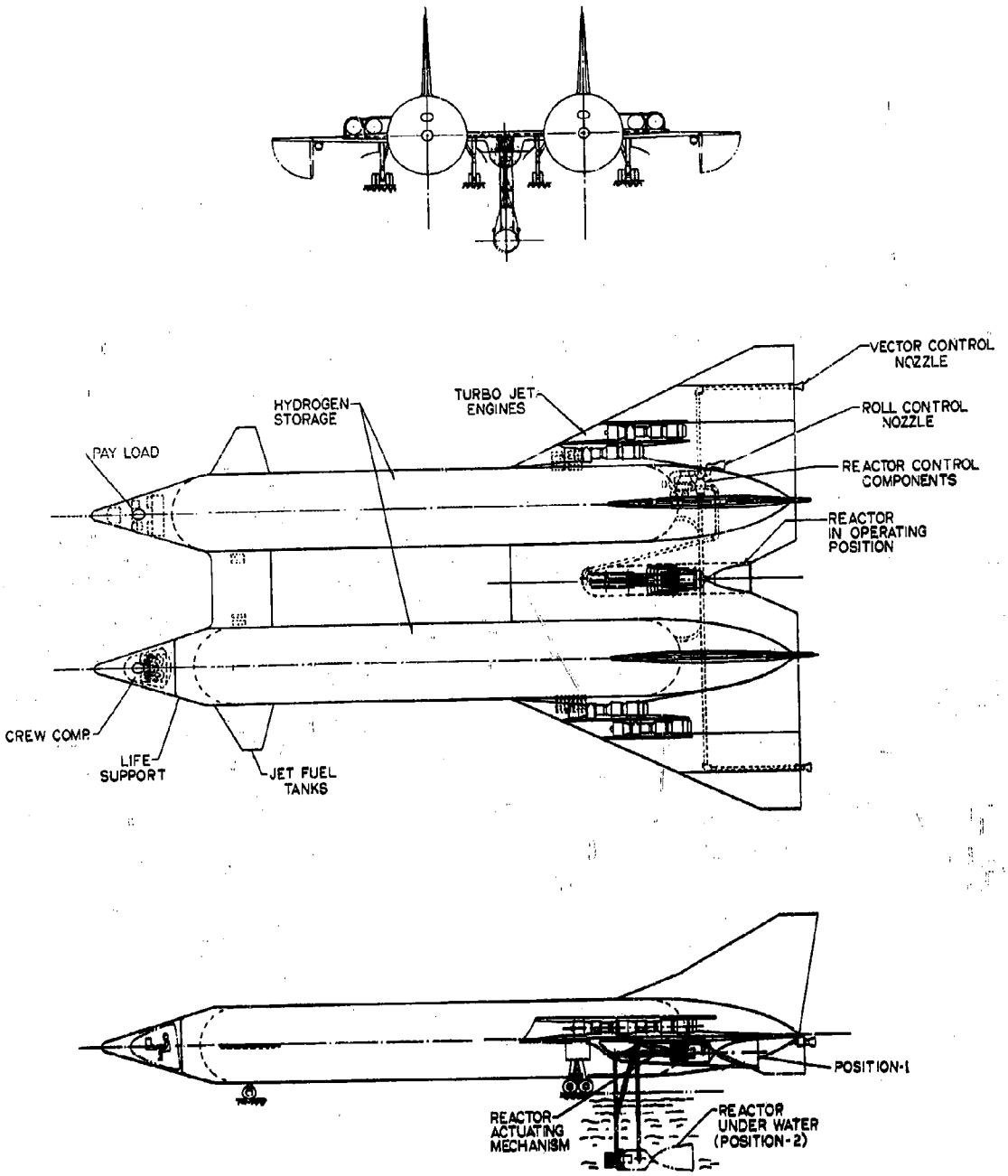


FIGURE 1. NUCLEAR AEROSPACE PLANE CONCEPTUAL DESIGN

Spaceship Maintenance

Based on our assumptions, there will be a requirement for a nuclear maintenance space craft which is sufficiently versatile to perform most any job involving the handling of radioactive components. We see such a craft as a very well outfitted self-propelled hot laboratory shop such as is shown in the conceptual drawing in Figure 2.

The principal items contained in such a craft are:

a. Shielded personnel space which serves as working and living quarters for the crew

b. Refueling equipment and spare propellant tankage

c. Welding, cutting and other tools to separate and assemble components

d. Spare parts

e. Propulsion and power plant

f. Decay heat exchanger, heat exchange fluid pumps and storage tanks

g. Shielded shuttle craft for personnel exchange

h. Communications

i. Docking system.

These maintenance ships will remain in space permanently and they will be supplied by aerospace freighters and by propellant supply ships.

Emergency Repairs

Minor repairs and maintenance, particularly on APU's should be made possible by operating crews without dependence on the service ship. Figure 3 shows a scheme where such components as pumps, heat exchangers, valves, and other items requiring frequent attention are located together behind a hollow shield wall which serves as a hot cell face. For repair, the secondary heat transfer fluid is pumped into the void space where it acts as a gamma shield and is at the same time protected from boiling into space when the system is opened. Repairs can now be carried out aboard the vessel without having to carry permanent shielding for such jobs.

NUCLEAR SYSTEM DESIGN FOR MAINTENANCE

The first requirement for successful maintenance can be stated simply: nuclear systems must be designed in basic modular units which can be serviced or replaced without tearing the whole system apart.

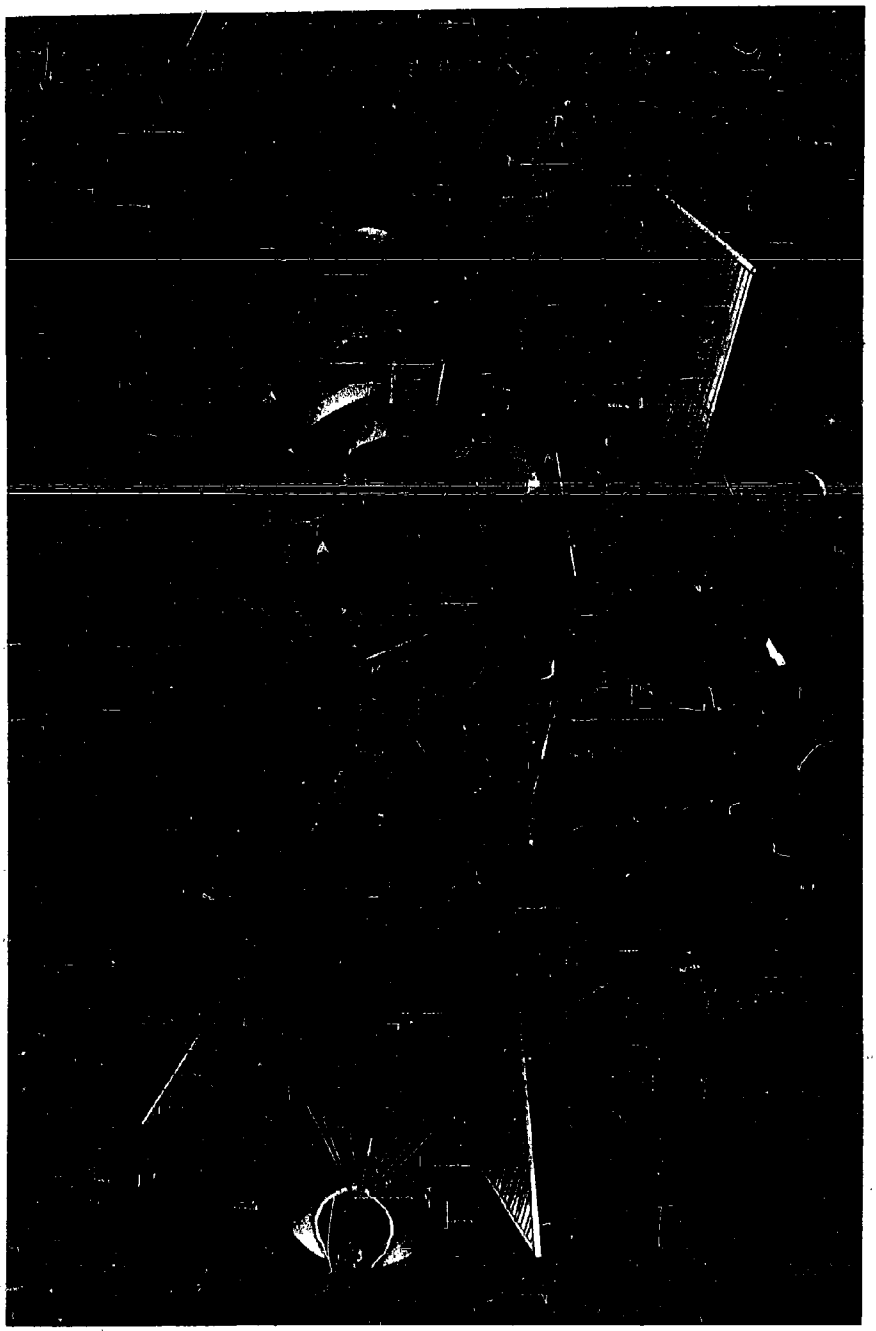


FIGURE 2. MAINTENANCE SPACESHIP

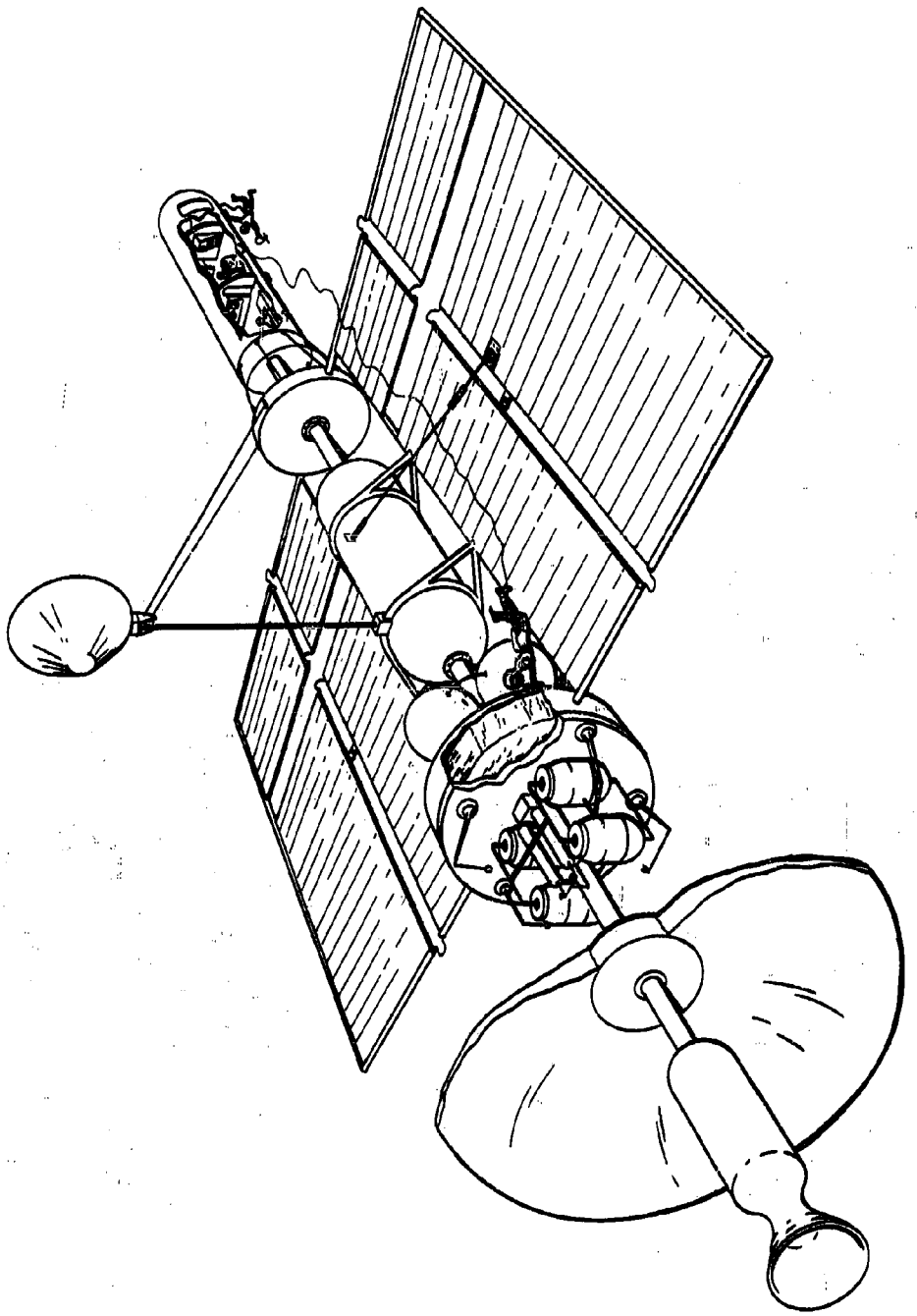


FIGURE 3. HOLLOW SHIELD REPAIR SYSTEM

Propellant Supply

Military craft may be virtually useless unless they can perform orbital changes or other maneuvers which make their instantaneous location difficult to predict. Propellant refueling is, therefore, vital. Regardless of the means by which propellants are collected and transported, nuclear systems can take advantage of the fact that they do not energize their propellant by oxidation but by direct heating. We should therefore be able to depend on whatever gas or mixture of gases is nearest or easiest to obtain. Hence, the rockets should be built so that they can use air, N_2, or any other gas rather than H_2 only. This will, of course, entail a loss in specific impulse, but for propulsion systems already in orbit this loss should be acceptable since they do not have to meet the performance requirements for terrestrial lift-off.

Moon or planet-based nuclear propulsion systems will similarly have to be fueled from whatever raw materials can be found locally. This will require the establishment of processing plants, but again, the reactor systems must be designed so that they can run off locally available propellants. In the event it proves to be impossible or uneconomical to build reactors which can first propel themselves to the planets and then run on locally available fuels, reactors built specifically for the latter fuels may have to be transported into space as dead weight.

Component Replacement

Figure 4 suggests a modular design in a schematic fashion with some attention to the frequency with which major components will require service.

A nuclear propulsion system might consist of six basic packages:

1. Propellant Supply
2. Nozzle
3. Core and Core Barrel
4. Reflector and Nuclear Control
5. Turbopump
6. Electronics.

There is little to be gained, at this time, by a discussion of specific means by which the nozzle or any other components will be removed or replaced. Suffice it to state that the designer will have to see to it that these parts are readily accessible in the order in which replacement will be required.

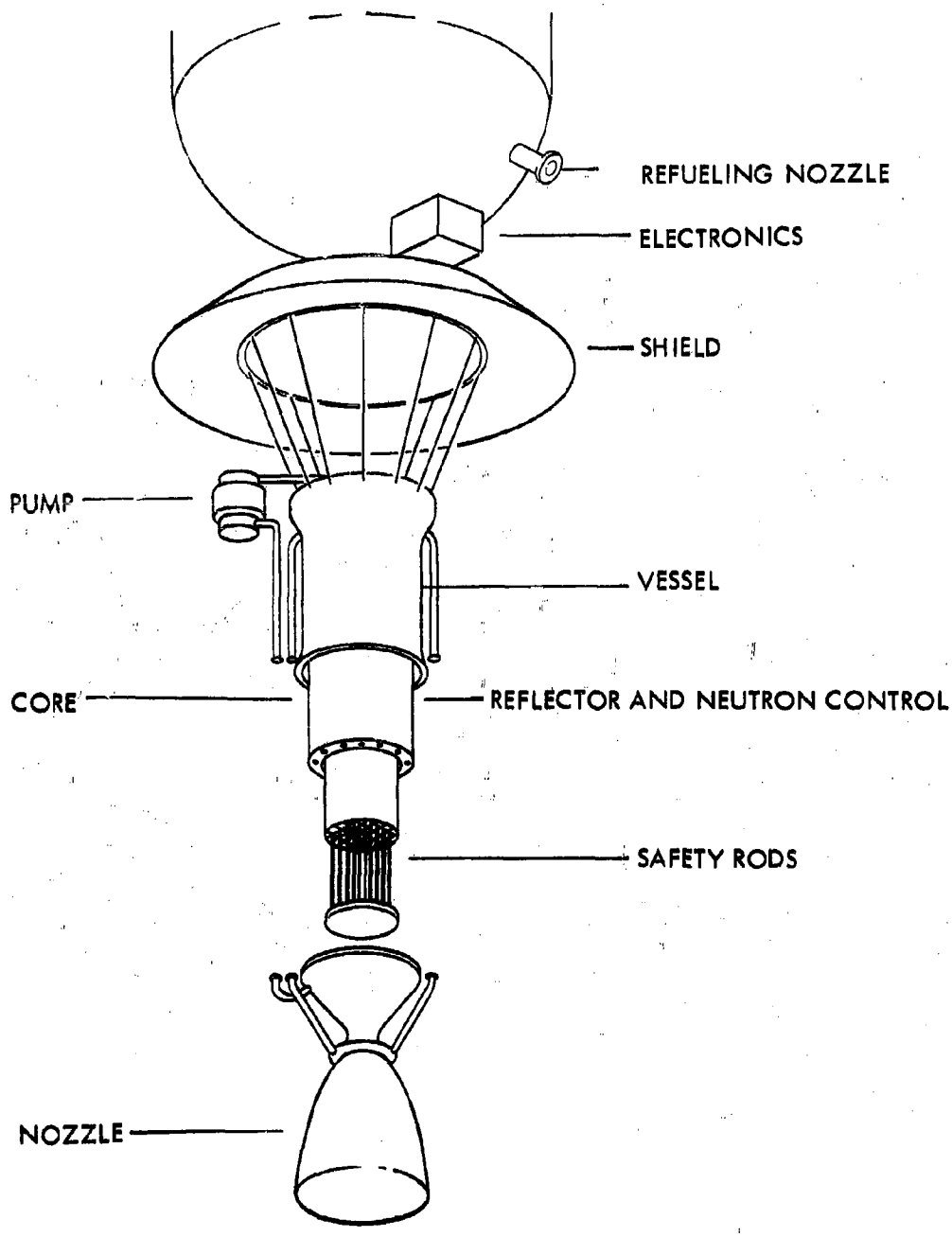


FIGURE 4. NUCLEAR PROPULSION SYSTEM MODULES

Propellant refueling frequency will exceed all other maintenance by orders of magnitude and must match present in-flight refueling systems in simplicity. The nozzle might go next, though there is evidence that if the core can stand the propellant flow at temperature, so will the nozzle. The core barrel could be held in place by the nozzle structure, so that core and core barrel can be replaced as a unit. Next, neutron poison controls and reflectors are removed as a cylindrical shell. Mechanical linkage of movable parts would mesh into control actuator linkage located within the reactor vessel. Both the turbopump and the electronic system can be replaced without disturbing any part of the reactor proper.

It would seem that core refueling is handled best by replacing the whole core. While burnup will undoubtedly not be uniform from fuel element to fuel element, fuel cycling or partial replacement of a core does not appear practical for military craft. Actual power plant refueling experience has shown that fuel cycling is difficult and time consuming as compared to total core replacement, since each individual element must be removed from its position in the core and placed into a specific, precalculated new position. In addition to the danger of damaging the elements during this operation, the replacement fuel inventory problem is complicated since the enrichment of the new fuel elements becomes a function of the operating history of the core. Any economic gains attributable to fuel cycling are thus likely to be offset by increased non-availability of the vehicle and inventory requirements.

Since cores of this type are relatively highly loaded, nuclear safety will play an important role in deciding just how such installations will be handled. As long as we are dealing with thermal reactors, temporary poison rods inserted into the fuel channels should provide adequate margin. The problems become more complex when dealing with fast reactors, where it might be necessary to replace the control system and the core simultaneously to assure that the core is never left without adequate nuclear control.

In view of the obvious problems which arise during solid core refueling, liquid fuel reactors should be re-examined for possible space applications. Both fused salts and liquid metal fuel systems are included here, the former being probably preferable due to the higher attainable fuel temperatures. This recommendation is made despite the acknowledged difficulties encountered in the development of such systems (2) since they would permit refueling by draining the depleted fuel and pumping fresh fuel into the reactor vessel.

APU and Power Station Maintenance

Refueling of space APU's should be simpler than refueling propulsion plants. Most of the designs, such as the SNAP-8 unit, contain the reactors in a vessel which has essentially one intake and one exit pipe for coolant flow. The core consists of individual fuel rods held in place by means of top and bottom support plates. Disassembly and maintenance of these units can be performed by the same equipment used for propulsion plant maintenance. Fuel cycling, by selective shifting of individual rods and the addition of a small number of new rods, is much more favorable here due to the simpler mechanical arrangement.

By adding auxiliary, normally closed, intake and exit pipes to the reactor vessels, servicing might be accomplished as follows. The auxiliary flow pipes are connected to a decay heat exchanger of the service vessel, and system flow pipes are cut. The whole vessel is then removed and a newly fueled reactor vessel moved into its place; the depleted core may now be refurbished at leisure.

The physical size of stationary space plants, regardless of their electrical output, should be restricted to that of the largest propulsion plant so that the same servicing equipment can be employed. The refueling methods used for present day nuclear power stations, with their tremendous water and auxiliary materials requirements, do not seem applicable. At the present time, and until radioactive material handling methods are improved by orders of magnitude, very large power requirements might better be met by using several interconnected smaller plants. The penalty of duplicating nuclear controls, reactor vessels and valving systems will be offset by increased reliability and the weight penalty may almost disappear for direct conversion systems.

SECONDARY USES AND DISPOSAL

Once nuclear space components have outlived their primary purpose, there arises the question of their disposal. The material was brought into space at some cost and it seems unforgivable waste to abandon it. We now encounter a well known lack of ideas on postirradiation use of materials. Most suggested applications run into hard economic facts even here on earth, where raw materials and equipment are relatively cheap.

Secondary Applications

One possibility that suggests itself derives from the fact that nuclear systems all require relatively large amounts of excess reactivity in order to override temperature defects, parasitic neutron absorption, etc. Even though such cores are depleted to the point where they no longer deliver design power, they can be useful in systems of different designs by changing the moderator. Thus SNAP-2 has 2% excess reactivity as designed and 7% excess reactivity in water. It should, therefore, be possible to remove a core from its vessel, place it into another vessel containing a hydrogenous coolant and extract additional power over long periods of time. True, the system may not be as efficient as one might wish since water is not the best medium for waste heat rejection into space; but let us assume a 5% efficiency in energy conversion. Figure 5 shows a Time vs Reactivity curve for the earlier mentioned graphite core, when used first for propulsion and subsequently in a pressurized closed cycle water plant. Assuming that the core can only be operated at 10% design power, it will still yield 5 MW for 540 days. Neglecting the materials question for the time being, reactor control would be the major problem to be faced. Chemical poison control by boron concentration in the coolant is an obvious solution; it may however not suffice in view of the extremely high δ k of the core. In that event a number of fuel elements might be removed and poison rods inserted in those positions.

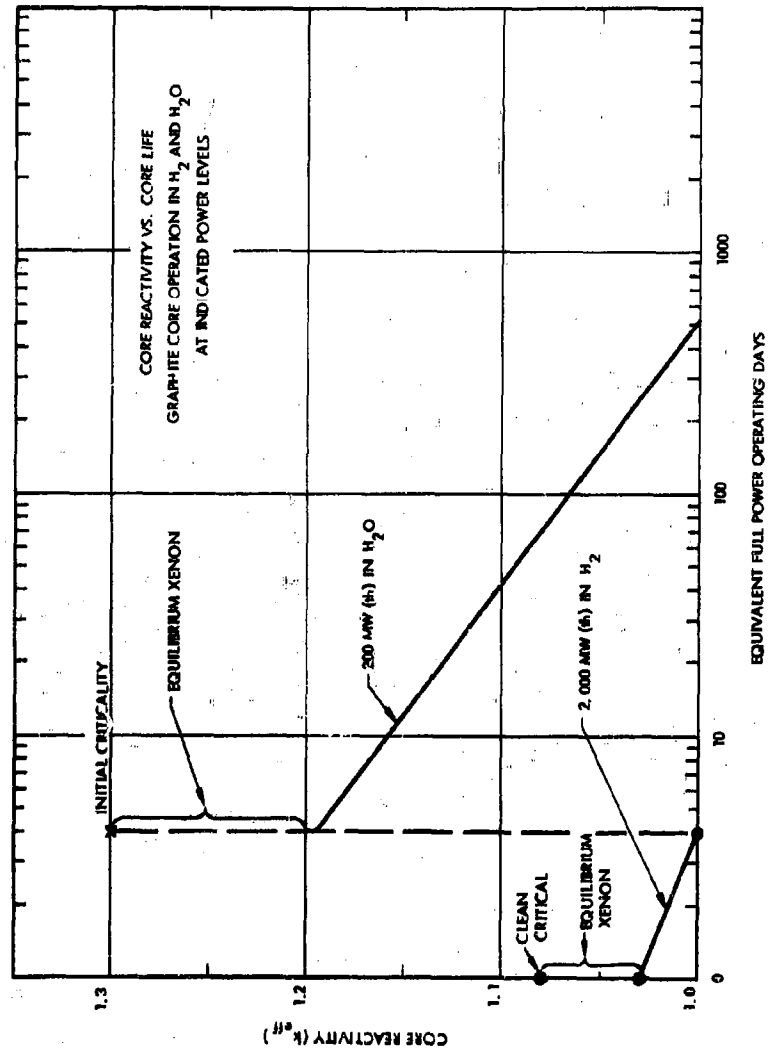


FIGURE 5. REACTIVITY VS. TIME IN HYDROGEN AND WATER

Subsequent to long irradiation periods, cores release considerable decay heat as shown in Figure 6 for the same core. This energy might be useful for space heating or for thermoelectric power generation in unattended locations as shown in Figure 7.

Other possible uses which seem attractive require isotope separation, Uranium reclamation, or extraction of "cold" materials, and had to be rejected since separation technology does not seem to lend itself to any practical system for the next few decades.

Scrap Disposal

Final disposal of radioactive scrap cannot be left to chance. Eventually, specific orbits or flight profiles will turn out to be more efficient than others and it is these that ships will travel and which would accumulate the scrap. The price of injecting such scrap into internationally agreed upon orbits, or ferrying it to a dumping ground on the moon might be compared to that of anchoring the mines which protect a harbor rather than leaving them free-floating. Even if these derelicts are equipped with warning devices powered by their own radioactivity, the orbital change required by a vehicle to avoid collision would be more expensive than orderly disposal.

Figure 8 illustrates the variation of orbital lifetime with initial orbital altitude assuming circular orbits. Accounting for the decay of fission product activity with time, the same figure also gives the relative level of fission product activity at the time of reentry. This shows that in the approximately 300 years it would take to spiral in from 550 n mi, the fission product activity remaining would be less than 1/10,000 of that present after only one year of decay.

For the assumed operation of 8,000 megawatt-days, after 300 years of decay only about 150 curies of total activity would be present; and about one curie of this would be Sr^{90}. At the end of 1,000 years, the Sr^{90} activity would be reduced by another factor of 10^{-8}. Other potentially dangerous isotopes show equivalent decay factors, and if disposal orbits of the type recommended in the next paragraph are established, earth contamination will not be a problem.

The Van Allen Belts are a second factor which affects the choice of disposal orbits. Refueling and maintenance would presumedly be handled in regions outside the belts to minimize radiation exposure to crews and systems. At the same time one might assume that orbits lying predominantly within the belts will be little traveled and might therefore make good disposal areas. Finally, one would want to expend as little energy as possible on the disposal process and establish the maintenance orbit as close to the disposal orbit as safety permits. Table 2 shows four suggested maintenance orbits and associated disposal orbits based on recent concepts of multiple Van Allen Belts. It is obvious that specific choices can be made based on strategic position and minimum disposal energy requirements.

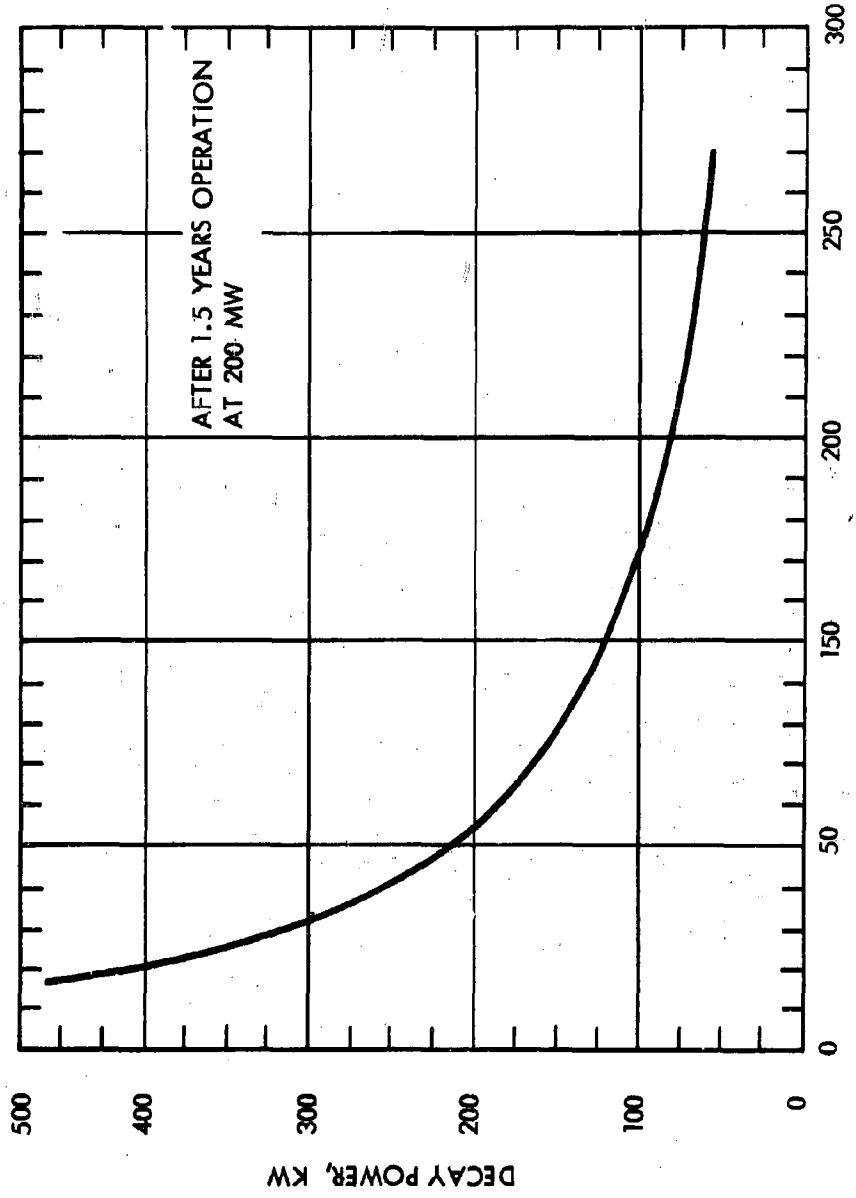


FIGURE 6. HEAT GENERATION DUE TO FISSION PRODUCT DECAY VS. TIME

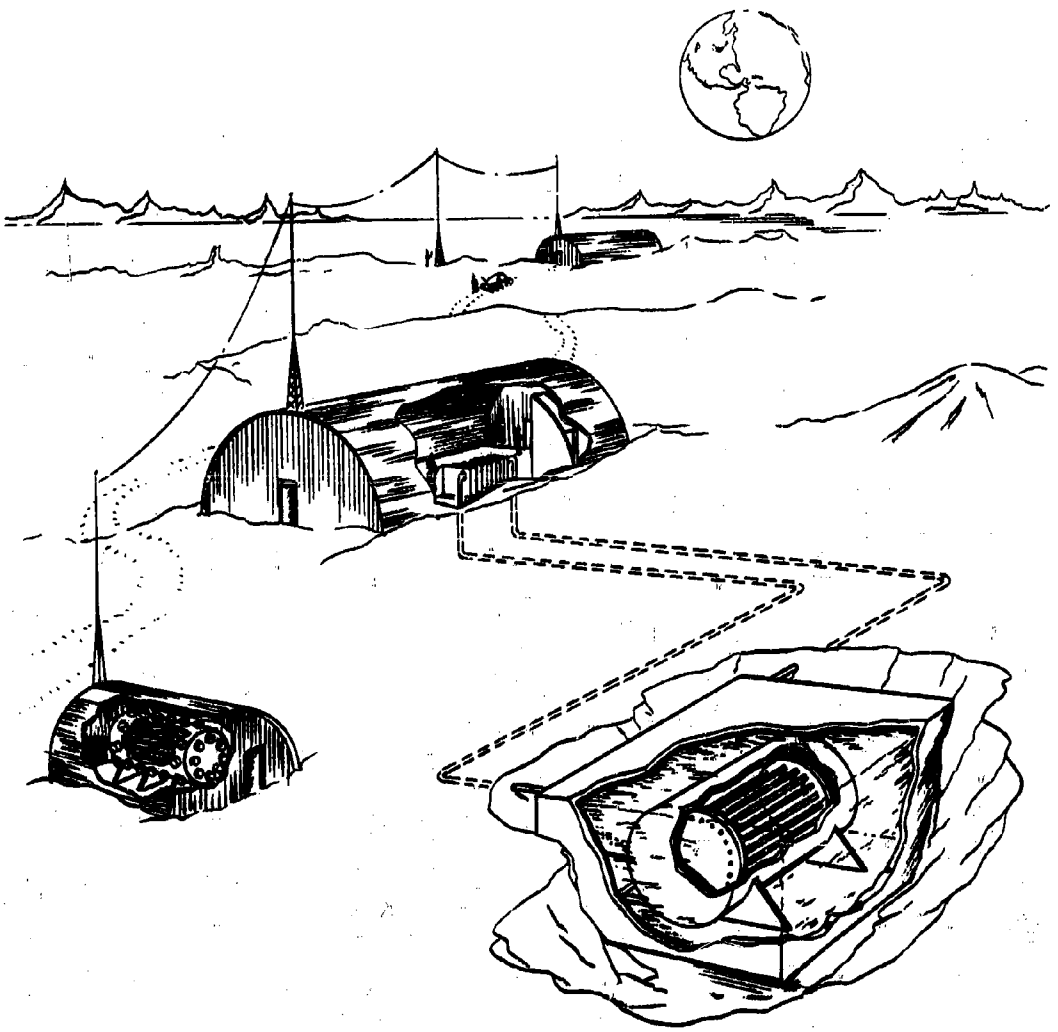


FIGURE 7. SPACE HEATING AND THERMAL ELECTRIC POWER

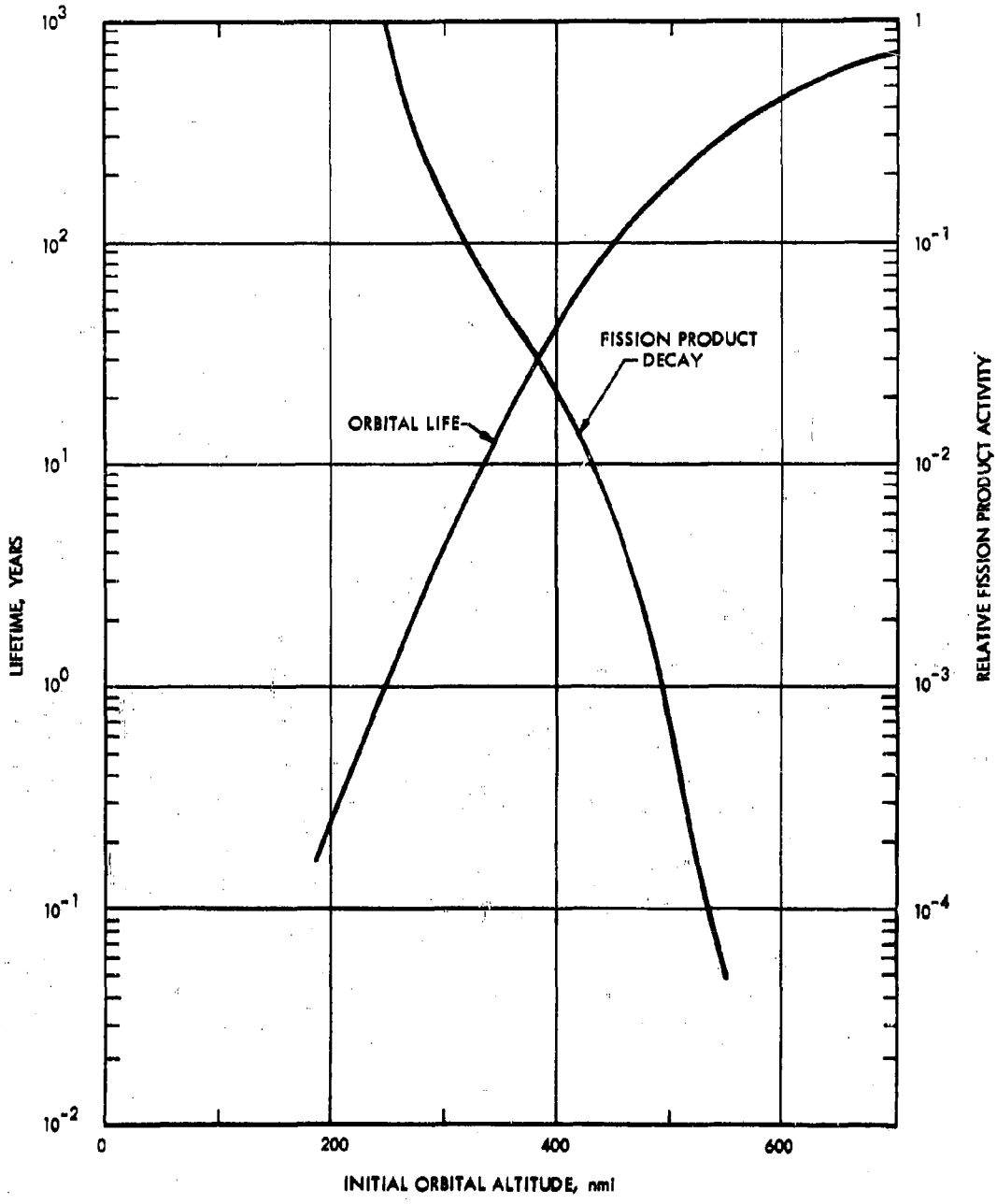


FIGURE 8. ORBITAL LIFE AND FISSION PRODUCT ACTIVITY VS. ORBITAL ALTITUDE

Table 2. Radioactive Disposal Orbits

Particles	Van Allen Belt* Energy (MEV)	Min. Orbital Altitude (Miles)	Max. Orbital Altitude (Miles)	Disposal Orbit (Miles)	Maintenance Orbit (Miles)	Maintenance to Disposal Orbit Δv Req'd (Miles/sec)
Protons	~100	1,400	2,800	1,860	940	0.36
Protons	20			7,500	6,200	0.16
Electrons	2	7,100	10,900			
Protons	0.25			1,000	11,500	0.11
Electrons	0.01	19,000	28,500	19,900	18,600	0.056
			47,000**	26,500	50,000	0.17

*Explorer XII Data (3)

**Max. Geomagnetic Field Boundary

In conclusion we might state:

1. Nuclear space systems for military applications must be designed from the start with the view of component replacement by remote maintenance.

2. Design of space maintenance ships for service of nuclear space systems should be initiated simultaneously with the design of the nuclear system itself.

3. Some mechanism must be found which will permit crews to make emergency repair on their own nuclear systems without imposing prohibitive shielding penalties on the vehicles.

4. Used radioactive components are not necessarily scrap but may be employed usefully in secondary applications.

5. Disposal of radioactive scrap in space must be controlled.

6. Interactions between solar or enemy generated radiation and nuclear space systems require further analysis.

The authors take this occasion to thank Dr. W. E. Shoupp for his help in the preparation of the paper, and Messrs. J. M. Weaver and R. F. Janz for performing many of the calculations.

REFERENCES

1. Koelle, H. H., "Economics of Planetary Travel Analyzed," Aviation Week and Space Technology, Vol. 76, No. 9, 7 May 1962, p. 59.

2. U. S. Atomic Energy Commission, "Report of the Fluid Fuel Reactors Task Force to the Division of Reactor Development," TID 8507, February, 1959.

3. Beller, W., "Explorer Finds Many Particle Belts," Missiles and Rockets, Vol. 10, No. 5, 29 January 1962, p. 31.

HYPERSONICS AND RE-ENTRY

AN EXPERIMENTAL EVALUATION OF SEVERAL ATTITUDE CONTROL CONCEPTS

Clarence J. Harris
General Electric Space Sciences Laboratory
Valley Forge, Pa.

and

Robert H. Johnson
General Electric Research Laboratory
Schenectady, New York

ABSTRACT

Several attitude control concepts have been considered for providing hypersonic re-entry vehicles with maneuvering capabilities. The attitude control concepts experimentally evaluated here involve the application of movable structures -- fins, flaps and rods -- located on the surface of axisymmetric ballistic-type re-entry vehicle models. Prior to considering these models with control surfaces, experimental studies were conducted to determine the static stability characteristics of and shock layer properties about the bodies without controls. Similar studies were then conducted on these bodies with controls present. Also evaluated was a magnetohydrodynamic concept for generating attitude control forces.

The experiments were performed in the General Electric Space Sciences Laboratory's 30" hypersonic shock tunnel located at Valley Forge, Pennsylvania, and in the General Electric Research Laboratory's helium tunnel located in Schenectady, New York. Data were obtained over a test section Mach number range of 4.7 to 28.9 and over a Reynolds number range of 1.3×10^4 to 0.5×10^6 per inch. Static stability data were obtained on centerbodies with and without control surfaces by using the shock tunnel free-flight test technique. Centerbody surface pressure distribution and flow visualization data were obtained both in the helium tunnel and the shock tunnel on sting-mounted models with and without control surfaces. MHD phenomena were studied in the shock tunnel on a sting-mounted two-dimensional model. The experimental results are compared to available and applicable analyses such as exact flow field calculations, Newtonian prediction, and viscous similarity correlations. These combined studies as

presented give a better understanding of the detailed centerbody control surface flow field phenomena involved in producing control forces. They further indicate the potentialities and limitations of the control devices studied.

Work done under U. S. A. F. Contract No. AF 04 (647)-617.

INTRODUCTION

A desirable capability for advanced generation hypersonic re-entry vehicles is the ability to maneuver along certain portions of their flight path. To achieve this, many flight attitude control concepts have been considered for ballistic-type bodies flying in the hypersonic regime. Among these is the concept of locating on the main centerbody movable external structures such as flaps, fins, rods, wings and spoilers. Another concept is to provide the centerbody with a variable center of gravity by means internal to the vehicle. Further, thrust devices such as jets and impulse rockets have been considered as well as concepts which utilize magnetohydrodynamically (MHD) induced phenomena to produce aerodynamic attitude control forces. It is the purpose of this paper to discuss experimental studies which were conducted to evaluate the feasibility of using several of these concepts.

The attitude control concepts discussed here involve movable surfaces on the main centerbody and MHD. The surfaces considered are flaps, fins and rods. These approaches are illustrated in Fig. 1 as applied to a typical re-entry shape. Shown is a slender, blunt, axisymmetric centerbody with:

1. movable fins located on the aft region of the centerbody. The fins illustrated are unswept, blunt leading edge wedges. Deflection of the fins produces pitch, yaw and roll control.

2. movable flaps hinged to the base of the centerbody. These flaps may be moved in and out of the centerbody flow field. The flaps may also be split (as shown) in order to achieve roll control in addition to the combined pitch and yaw control.

3. movable rods located at various centerbody stations. These rods are capable of being projected varying depths into the centerbody shock layer. It may also be desirable to spin or vibrate the rod or rods in order to reduce the local aerodynamic heating to these surfaces. As illustrated here a single rod device may be used to produce variable pitch control.

and

4. an MHD two-dimensional channel located in a slot arrangement in the aft region of the re-entry body. The centerbody contains an electromagnet located below the channel. This magnet may be energized by a separate internal power supply or by the MHD induced electrical power extracted from the channel by electrodes, such as the segmented electrode shown in the illustration. This concept utilizes

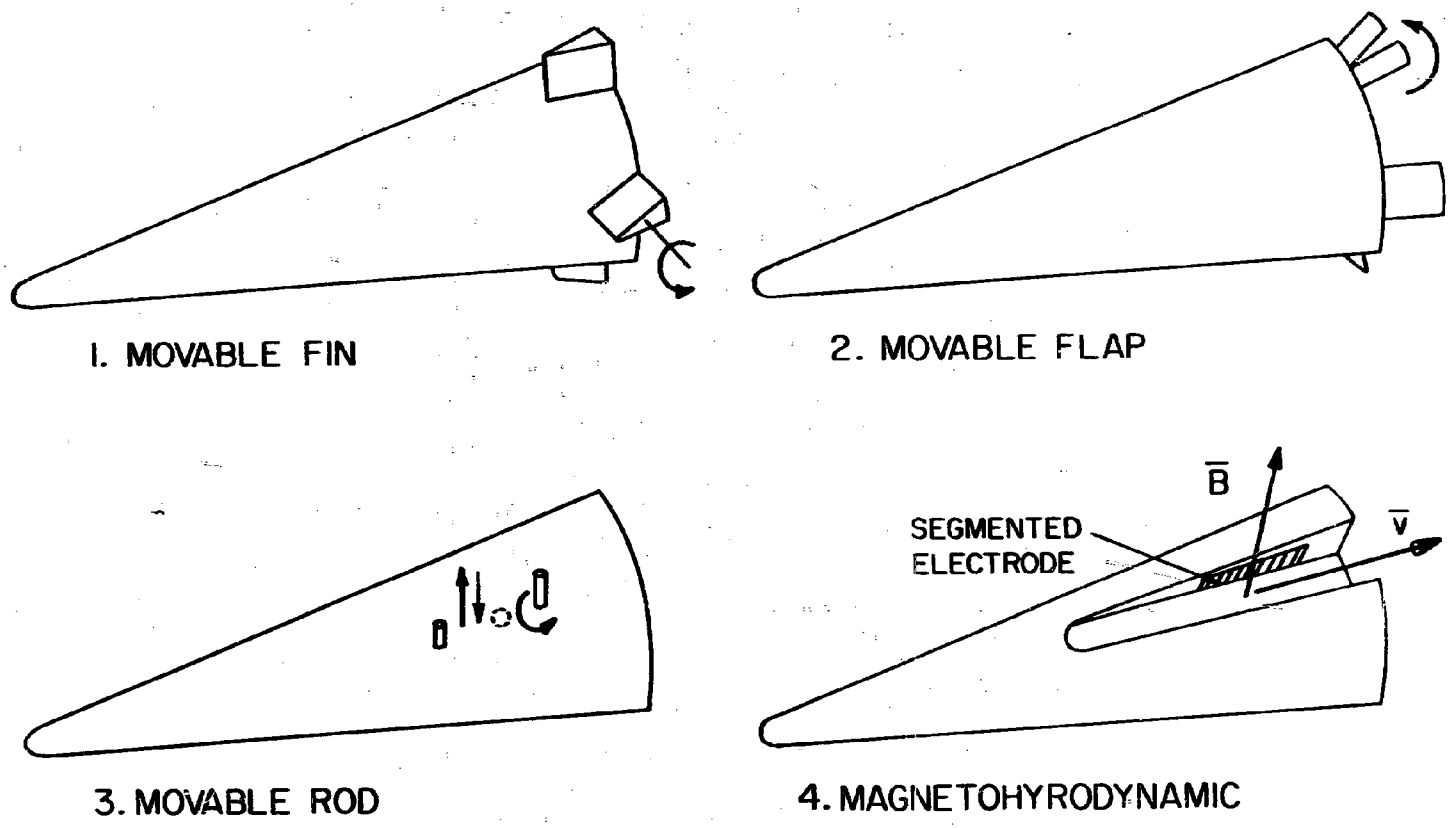


Fig. 1. Illustrative Ballistic Missile Attitude Control Concepts

the induced Lorentz force resulting from the flow of ionized shock layer gases through the magnetic field to achieve control. It may be desirable in this concept to seed the shock layer flow in order to increase its electrical conductivity and therefore the magnitude of the control force.

These experimental studies involved two investigations. One was the acquisition of data which would show the gross modifications of the bare centerbody's static aerodynamic characteristics due to the presence of fin, rod and flap control surfaces. This consisted of obtaining static stability data, i.e., C_D, C_L, L/D, and Xc.p. measurements as a function of both centerbody angle of attack and some particular control parameter, δ Fig. 2 presents illustrative static stability data. Control effectiveness is considered here to be the extent to which L/D varies (i.e., the range varies), due to a given variation in the control parameter, δ at some initial α value (α_o). This may be evaluated in the following manner using the illustrative experimental data. An arbitrary design center of gravity is first defined (c.f. Fig. 2) for the body with attitude controls. In order that the body have zero moment (stable trim position) it is necessary that the center of pressure and center of gravity coincide. The dotted line of Fig. 2 shows the variation in L/D with α to be expected when the foregoing condition is satisfied. It is seen that under these conditions, due to varying the control parameter from 1 to 2, that the L/D ratio increases, thereby extending the horizontal or lateral range capabilities (depending upon the frame of reference) of the vehicle. A decrease in the range is caused by control perturbations in the regime above 2 or below 1. The dash-dot lines (Fig. 2) are static stability characteristics for the centerbody without controls. The body alone center of pressure characteristic curve is typical of that found for slightly blunted, slender axisymmetric bodies. As such, this type of Xc.p./ℓ vs α dependence results in a statically unstable configuration. The addition of control surfaces to this centerbody may be expected, as illustrated here, to produce the desired positive Xc.p./ℓ vs α slope needed for a statically stable configuration. At present, the ability to define the static stability characteristics for a body with control surfaces is limited. This is due in part to limitations in the currently available analytical techniques which are used in defining the resultant forces on blunt centerbodies which have shock layer flow fields perturbed by complex centerbody-control surface interactions.

The second investigation deals with this interaction problem. Here the details of the flow field perturbations which result from centerbody-control surface interactions were studied using pressure measurements, Schlieren, and boundary layer-oil flow techniques. The results of these measurements may be used in a phenomenological manner to predict the mode by which control forces are generated by flap, fin and rod surfaces.

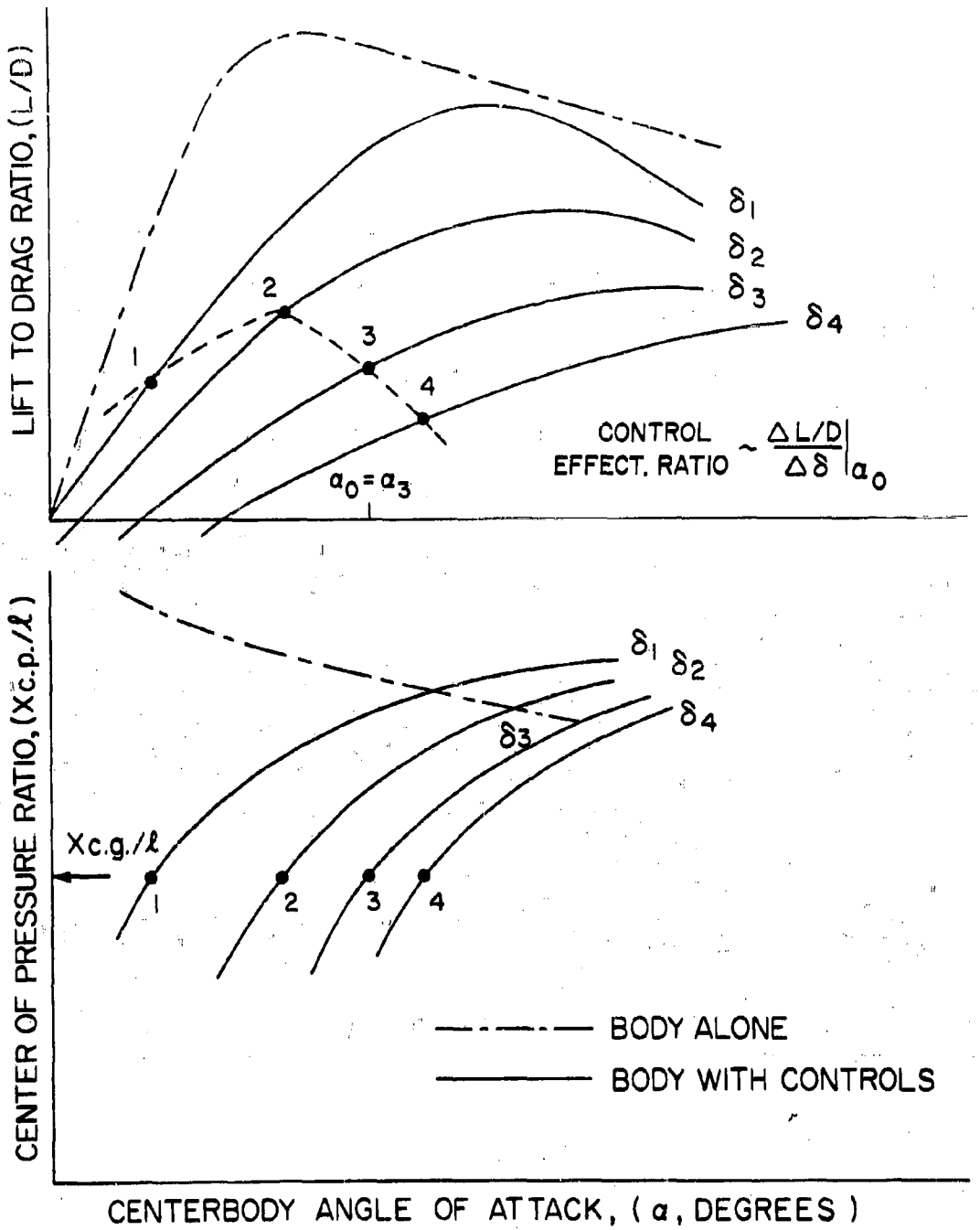


Fig. 2. Illustrative Static Stability Characteristics

The dependence of these interactions on free-stream Mach number and Reynolds number is also discussed.

The generation of control forces by an MHD channel, of course, involves more than just the flow interactions of the previous three concepts. The MHD interactions may be expected to be quite complex. Very little experimental work is available which clearly defines the extent to which first-order theoretical MHD relationships will apply in a realistic hypersonic flight attitude control problem. For this reason the experimental study in this area was of a somewhat basic nature and less applied than the flap-fin-rod studies. The primary objective was to design and test a workable MHD two-dimensional channel under shock layer flow conditions which would simulate those to be expected in flight on a typical axisymmetric re-entry vehicle. The results of this test were then evaluated in terms of available theories.

EXPERIMENTAL PROCEDURES

Considerable experimental static stability data have been obtained in the laboratory on fin-controlled re-entry shapes (1). In these earlier studies the centerbody was a rather blunt, medium drag, 9° sphere-cone. These data were obtained in air in the G. E. 30" shock tunnel primarily at a free-stream Mach number of 12.6 and at a free-stream Reynolds number per inch of 1.39×10^5. In shock tunnel testing the free-flight technique was used exclusively in acquiring static stability data (2) (3). A Schlieren photograph of a free-flying fin-controlled re-entry vehicle is shown in Fig. 3. The semi-circular bulge appearing in the figure is the image of the mirror in the optical system. Other types of data were also obtained on this configuration. These included flow visualization, pressure and heat transfer distribution measurements obtained both in the 30" shock tunnel (4) and the helium tunnel (5) over a wide range of hypersonic free-stream test conditions. The combined shock tunnel-helium tunnel test data determined the effects in these studies of such fin variables as fin deflection angle, fin span, fin leading edge sweep and fin leading edge thickness. These data were then utilized in predicting the performance of various finned-controlled re-entry vehicles over typical flight paths (1). These earlier data therefore provide a yardstick for evaluating the present data which were obtained using similar and other control concepts and centerbody shapes. This comparative approach as utilized in the current study (the entire study is summarized in Table 1) involved acquiring experimental static stability data on a 0.3 bluntness, 9° sphere-cone having both flap and rod attitude controls. These data were obtained at the $M_\infty = 12.6$ shock tunnel free-stream test condition. Pressure, boundary layer flow and schlieren data were also obtained but on a 0.12 bluntness, 9° sphere-cone with a flap control. The schlieren and boundary layer flow

SIMULATED HYPERSONIC CONTROLLED RE-ENTRY

$M_\infty = 12.6$

Fig. 3. Shock Tunnel Free-Flight Model (Schlieren Photograph)

Table 1

Facility	Measurement	Free-Stream Conditions			Configuration
		Medium	M_∞	Re_∞/in.	
30" Shock Tunnel	C_L, C_D and Xc. p. /l	Air	12.6	1.39×10^5	9° cone with controls ($r/R = .3$)
Helium Tunnel	Surface Pressure & Flow Visualization	Helium	21.3	5×10^5	9° cone with controls ($r/R = .12$)
30" Shock Tunnel	Surface Pressure	Air	12.0	1.5×10^5	9° cone with controls ($r/R = .12$)
30" Shock Tunnel	C_D ($\alpha = 0°$)	Air	17.5	1.3×10^4	9° cone, 3/4 power law ($r/R = 0, .05, .12$)
30" Shock Tunnel	C_L, C_D and Xc. p. /l	Air	17.5	1.3×10^4	3/4 power law & 9° sphere-cone without control ($r/R = .05$) 3/4 power law with flap ($r/R = .05$)
30" Shock Tunnel	MHD	Air (seeded)	4.8	1.3×10^4	2-dimensional wedge with MHD channel

data were obtained in the helium tunnel at $M_\infty \sim 22$.

Also investigated was the effectiveness of a flap in achieving attitude control for a centerbody with zero lift drag characteristics lower than the previous 0.3 bluntness, 9° sphere-cone body. Static stability data at $M_\infty = 17.5$ and Re_∞/in. $= 1.3 \times 10^4$ were obtained on several low drag bodies without control surfaces. These configurations are shown in Fig. 4. They consisted of a 3/4 power-law body, and a 9° cone. Also shown is a tangent ogive body. Zero lift drag data were obtained for these three body shapes as a function of the nose bluntnesses shown (0, 0.05 and 0.12) (6). The 0.05 nose bluntness, 3/4 power law body was then tested at $M_\infty = 17.5$ with flap control. Bare body angle of attack static stability data were also obtained on the 0.05 bluntness sphere-cone and 3/4 power law bodies. The dimensions of the control surfaces are given in Table 2.

The MHD study was performed in air in the test section of the 30" shock tunnel. The model used in this study is shown in Fig. 5. This body is a hemicylindrically blunted 17° half angle wedge on which is mounted a two-dimensional MHD channel containing removable electrodes. Under the channel is an air-core electromagnet. In the experimental study the magnet was externally powered. The model was mounted on a drag balance (7) attached to the sting support system. Provisions were also available for measuring the surface pressure on the channel floor and for measuring the induced MHD electrical power characteristics under varying external resistive load (R_L) conditions. The attempt was made during this experiment to optimize the overall testing conditions (8) by 1, properly choosing model design parameters and test section flow characteristics, 2, seeding the model shock layer flow field with a potassium salt, and, 3, heating the channel electrodes. As shown in Fig. 5 the first-order MHD relationship desired is the generation of the Lorentz force ($\bar{F}$) resulting from the flow of the shock layer gas with a velocity $\bar{v}$ in the presence of the channel magnetic field $\bar{B}$. The channel when located on just one side of the body produces an unbalanced force which provides the means for achieving attitude control.

As stated the tests were performed both in the 30" shock tunnel at the G. E. Space Sciences Laboratory, and the helium tunnel at the G. E. Research Laboratory. The 30" shock tunnel is shown in Fig. 6. The performance and operation of this facility are detailed in references 9 and 10. Fig. 7 shows the helium tunnel. The performance and operation of this facility are detailed in reference 11 .

Analyses of the experimental data were attempted in terms of available predictions, theories and correlations. The zero lift drag static stability data are analyzed in terms of viscous aerodynamic similarity parameters (12, 13, 14, 15). The result obtained from using this type of

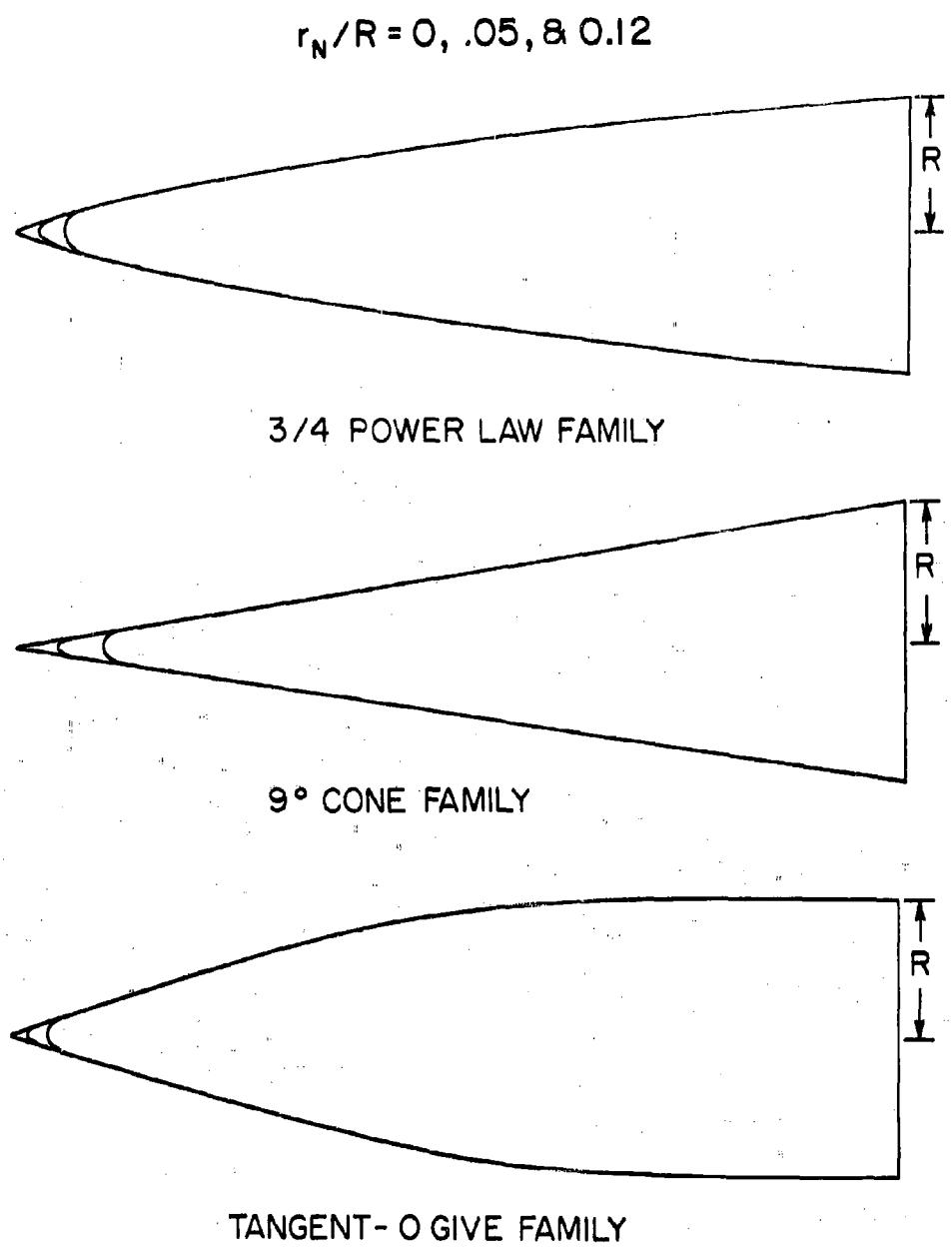


Fig. 4. Re-entry Shapes with Various Nose Bluntnesses

Table 2. Control Surfaces

Type	Design	Location on Body
Flap	Shape: square	Base of centerbody
	Area: 15% of centerbody base area	
Fin	Shape: hemicylindrical, wedge (for details see ref. 1, Fig. 2)	Aft end of centerbody, in a cruciform array
Rod	Diameter: 0.055 of the centerbody base diameter	X/D = 2.0 station on centerbody
	Height: 0.17 of the centerbody base diameter	

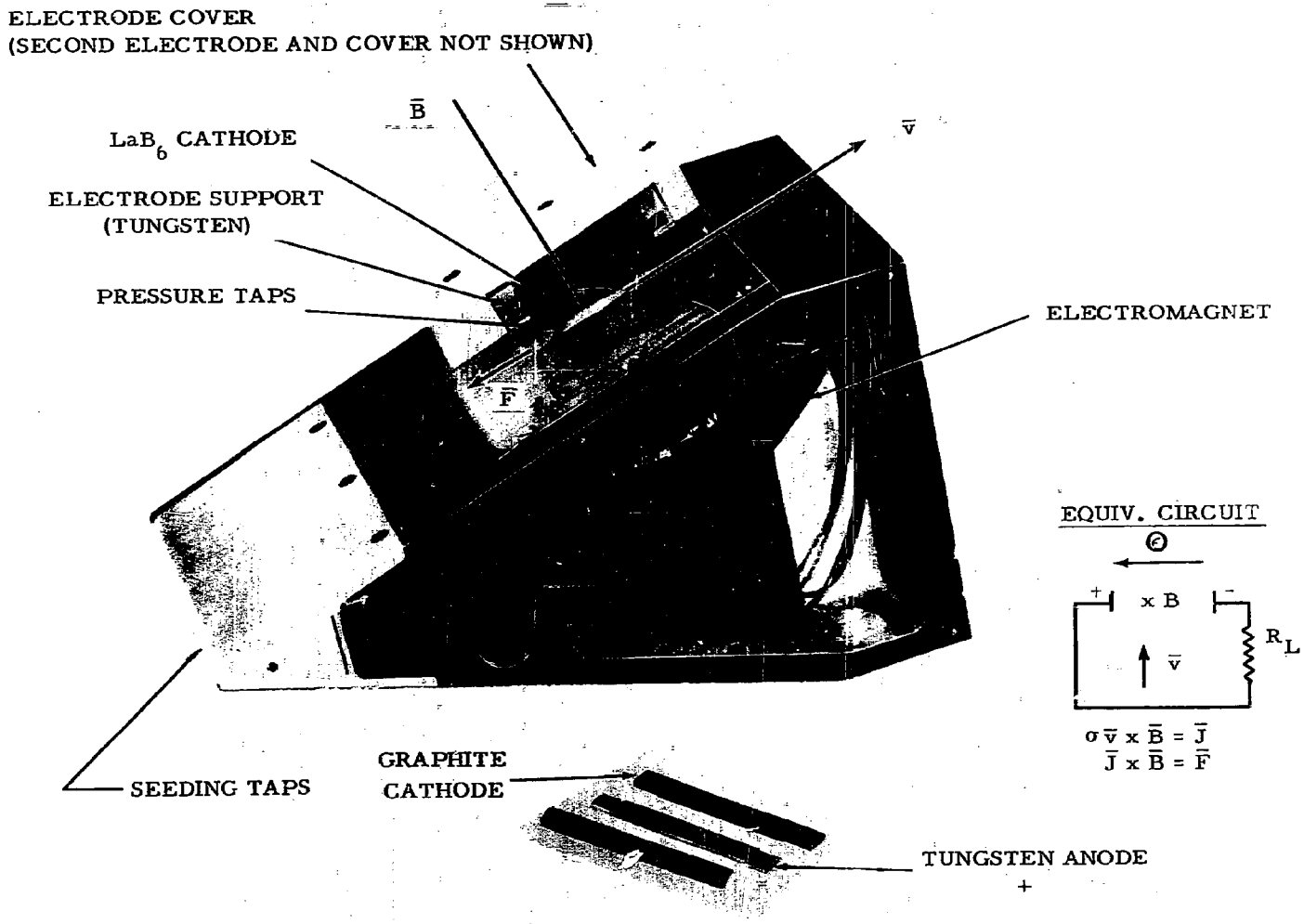

ELECTRODE COVER
(SECOND ELECTRODE AND COVER NOT SHOWN)
B
v
LaB6 CATHODE
ELECTRODE SUPPORT
(TUNGSTEN)
PRESSURE TAPS
ELECTROMAGNET
F
EQUIV. CIRCUIT
x B
RL
v
σ v x B = J
J x B = F
SEEDING TAPS
GRAPHITE
CATHODE
TUNGSTEN ANODE
+

Figure 5

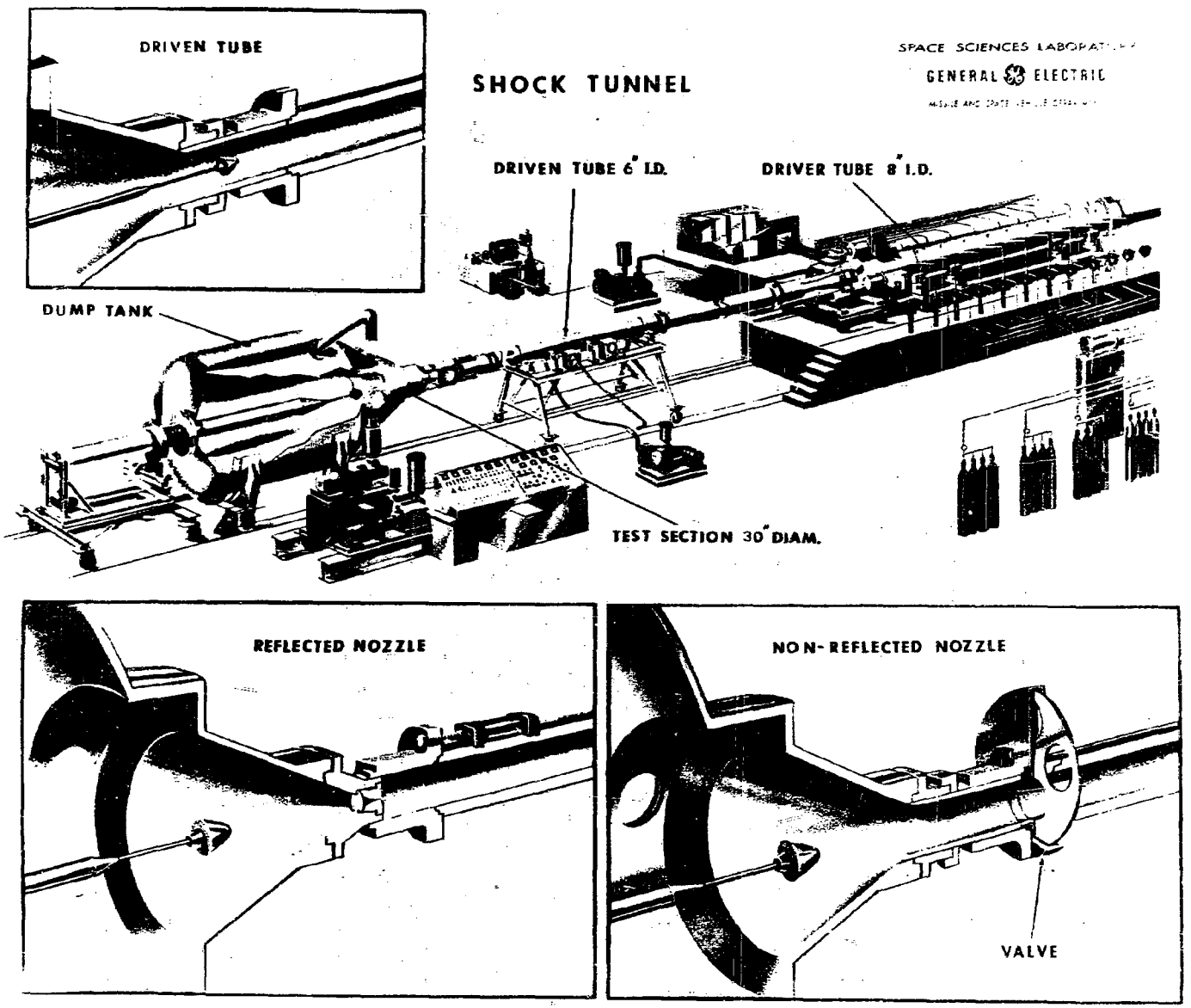

DRIVEN TUBE
SHOCK TUNNEL
SPACE SCIENCES LABORATORY
GENERAL ELECTRIC
DRIVEN TUBE 6" I.D.
DRIVER TUBE 8" I.D.
DUMP TANK
TEST SECTION 30" DIAM.
REFLECTED NOZZLE
NON-REFLECTED NOZZLE
VALVE

Figure 6

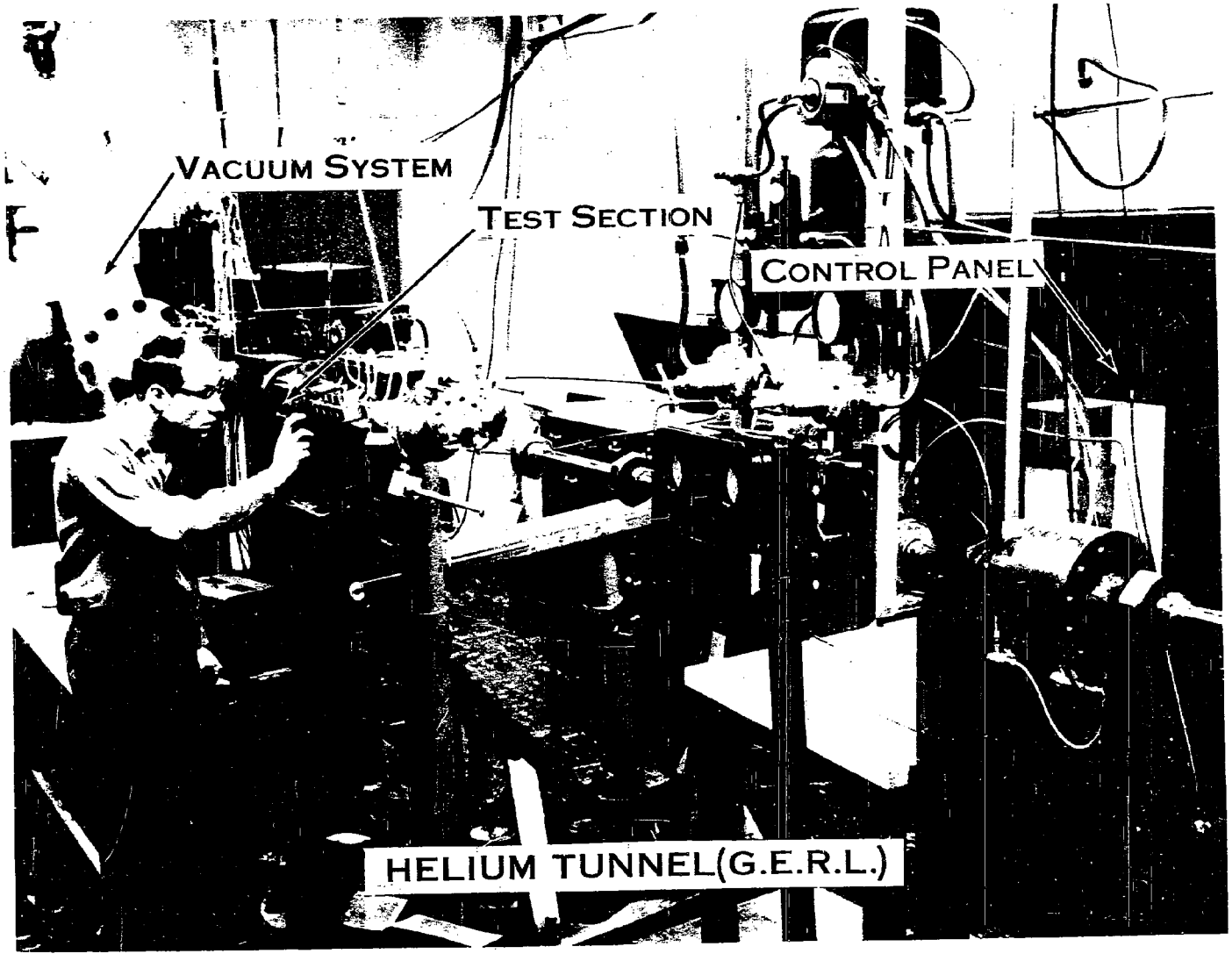

VACUUM SYSTEM
TEST SECTION
CONTROL PANEL
HELIUM TUNNEL(G.E.R.L.)

Figure 7

an analysis was utilized as an approximate correction to the Newtonian L/D vs α prediction (16) and the corrected prediction is compared to the obtained data.

The bare body pressure data are compared both to the pressure correlations offered by Cheng (17) for slightly blunt bodies and to exact flow field solutions (18, 19). The control surface influence on axisymmetric centerbody boundary layer flow separation is empirically analyzed with the aid of two-dimensional separation data obtained in helium at $M_\infty \sim 28.9$ (20). The MHD data are compared to fundamental analytical results such as offered by Cowling (21), Sears (22), and Resler (23).

RESULTS

Aerodynamic Controls

The experimental data showing the comparative performance of a blunt, axisymmetric re-entry shape using flap, fin or rod attitude control are given in Fig. 8. The centerbody was a 9° sphere-cone with $r/R = 0.3$. The nominal model length is 3". The bare body lift to drag ratio and center of pressure characteristics are also presented. The L/D data for the bare body are found to fall well below the Newtonian prediction. However, Geiger (16) has suggested an approximate viscous drag correction to the inviscid Newtonian prediction which when applied results in far better agreement between the theory and experiment. The addition of control surfaces reduces the L/D performance. A fin control system composed of four unswept leading edge fins in a cruciform array at the aft end of the centerbody produces unstable and therefore undesirable static stability characteristics above a centerbody angle of attack of 8°. The pitch fins were deflected -30°. The dorsal fins were undeflected. The ability to obtain more desirable static stability characteristics with a fin system is discussed in detail in reference 1. A single flap located on the low pressure side of the body and having a deflection angle of 30° with the centerbody axis produces statically stable center of pressure characteristics. This also occurs with the single rod attitude control element. The rod is located at the centerbody's $X/\ell = .85$ station and is perpendicular to the centerbody axis. The center of pressure results indicate that the flap control generates a small negative normal force component which acts at an axial body station some distance upstream of the aft end of the body but aft of the body along Xc. p. $/\ell$ location. The rod control appears to generate a similar but smaller force component which acts further back on the body, probably close to the region where the rod is physically located. Thus, it appears (as expected) that the flap influences a relatively large area of the centerbody surface upstream of its actual location on the centerbody.

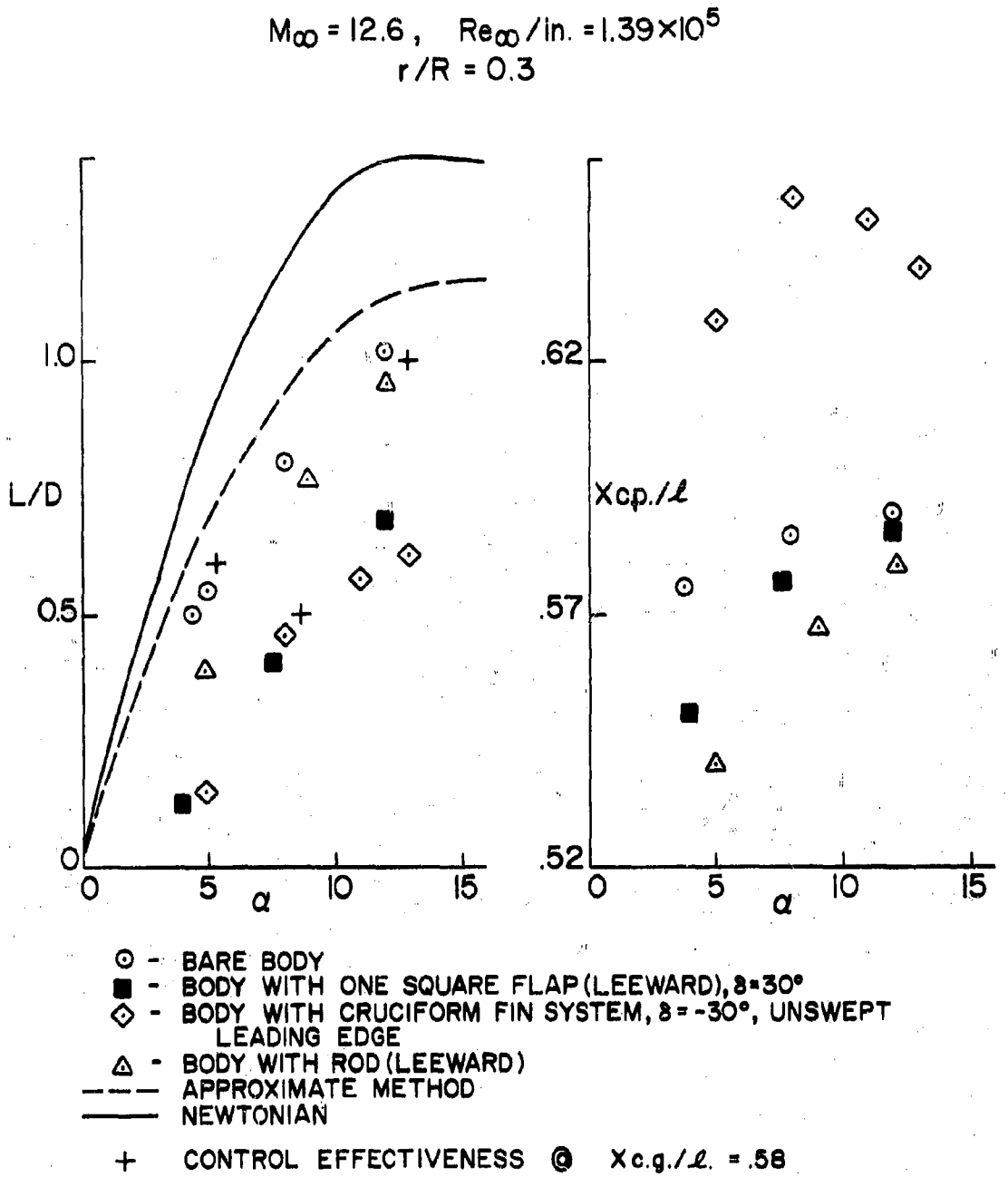


Fig. 8. Comparative Flap, Fin, and Rod Static Stability Performance, 9° Sphere-Cone

In the earlier fin attitude control study (1) the experimentally determined static stability characteristics were used in an IBM 7090 computer program to obtain flight trajectories for several typical fin-controlled re-entry shapes. One centerbody was the 9° sphere-cone with $r/R = 0.3$. The computed trajectories for this centerbody show that an appreciable increase in lateral range over the zero lift finned body value could be obtained with the unswept fin control system when the statically stable portion of its angle of attack (l/D) characteristics were utilized. Assuming flight in the higher angle of attack range ($\alpha \sim 20°$ and $L/D \sim 0.8$), a further increase in range was predicted but, as stated before, this is the angle of attack range where the control configuration is statically unstable. With flap control the experimental results indicate that a larger range increase may be obtained ($\alpha \sim 15°$) and under statically stable flight conditions (the static margin will however be small). Similar performance may be expected with the rod control concept with the possibility that even a greater increase in range and static margin may be obtained.

The flap and rod data as presented does not show control effectiveness per se since only one control parameter (δ) was studied for each control concept. Detailed control effectiveness data were obtained previously on fin control (1). However, an indication of the control effectiveness due to the flap or rod may be obtained by considering the measured aerodynamic characteristics for the body with and without these controls. These results are shown by the crosses in Fig. 8. Choosing an $X_{c.g.}/\ell = .58$, the L/D values at the stable trim conditions may then be defined for the bare body, the flap body, and the rod body. The flap reduces the L/D value by approximately 17% and the rod increases the L/D value by approximately 67%. The change in range to be expected when going from the bare body zero lift condition (the ballistic trajectory) to the flap or rod control condition may also be estimated in the foregoing manner.

The preceding data are of course gross performance data. The criteria for comparison were the static stability characteristics as obtained for the same centerbody re-entry shape with different control concepts at the same free-stream test conditions. No effort was made here to evaluate the performance in terms of such comparative control surface parameters as control surface area. However, the details of how these control surfaces influence the centerbody shock layer flow field and thus generate control forces due both to the resulting induced control surface pressures and centerbody surface pressures were considered and experimentally studied. Fig. 9 presents centerbody surface pressure distribution data obtained in air and helium on 9° sphere-cones at zero angle of attack without control surfaces. The air data were obtained at $M_\infty = 12.0$ and $Re_\infty/in. = 1.5 \times 10^5$ on a $r/R = 0.3$, 9° sphere-cone body ($\ell \sim 6''$). The air and helium data agree and further are in very good agreement with

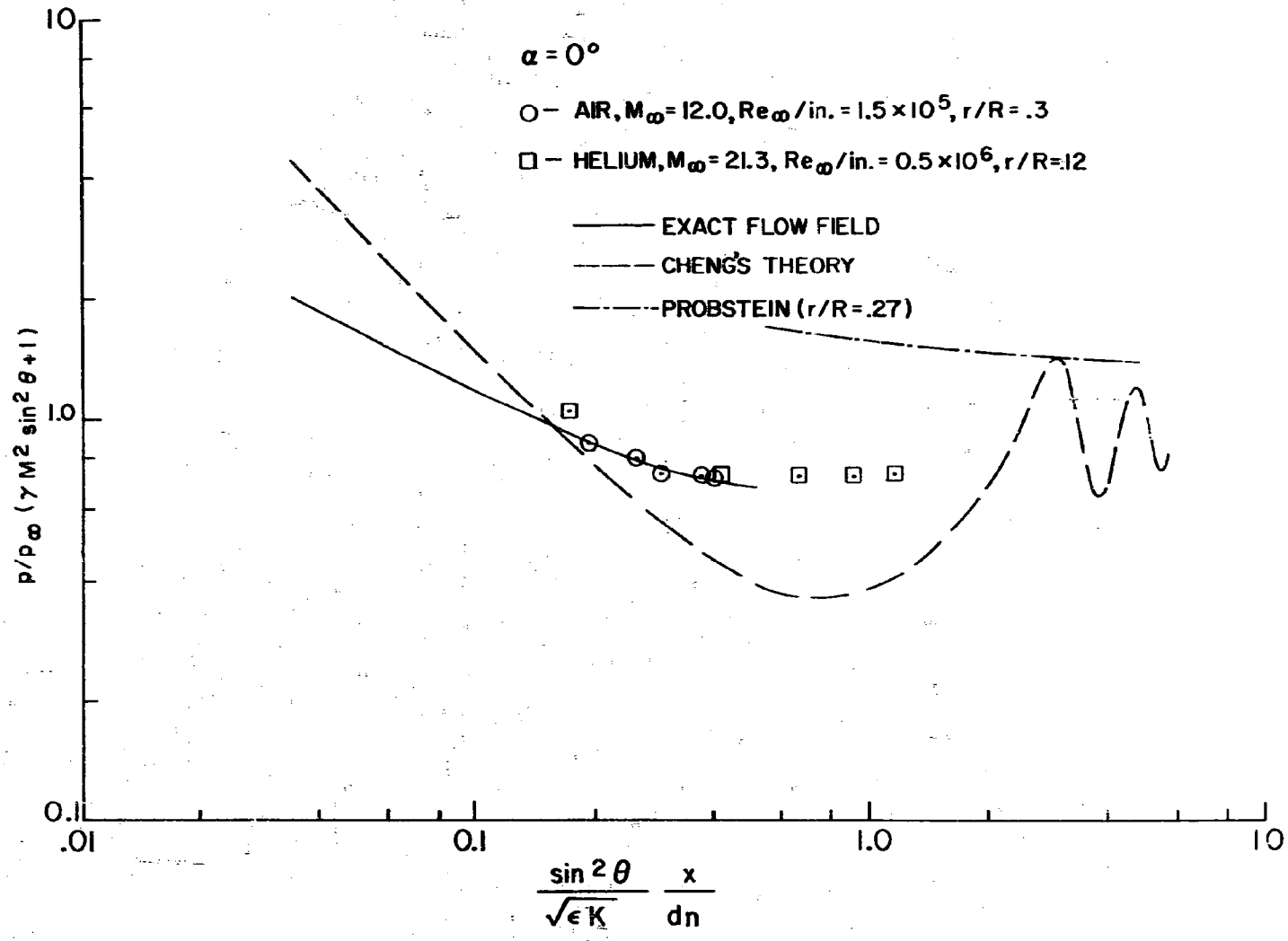


Fig. 9. Pressure Distribution on 9° Sphere Cone (Helium and Air)

exact flow field predictions (18) (19) but lie above the prediction of Cheng (17) and below the purely viscous prediction of Probstein (13).

The degree to which the presence of an attitude control surface affects the centerbody surface pressure is shown in Fig. 10 for a flap body. Both air and helium data are presented. At $\alpha = 5^\circ$ in the helium test there is a negligible increase in centerbody surface pressure. The pressure measured on the flap surface is presented in terms of ψ, the ratio of the measured flap pressure to the pressure measured at the rearward-most centerbody station without a flap present. The very slight rise in pressure which occurs on the deflected flap precludes the formation of an oblique shock in this region. In the air study, also at $\alpha = 5^\circ$ but $\delta = 90^\circ$, centerbody surface pressure increases as high as a factor of 17 were obtained. The average pressure increase was, however, but a factor of 3. At $\alpha = 0^\circ$ for the helium test the flap protrudes through the separated boundary layer into the shock layer and an increase in centerbody surface pressure was measured. The flap surface pressures measured indicate that a strong shock exists near the tip of the flap. Extensive pressure and boundary layer flow measurements have been previously obtained with fin controls at M_∞ = 12.0 on bodies tested in the shock tunnel (1) and at M_∞ = 25.9 in the helium tunnel (5). These measurements showed that large perturbations occur in the centerbody shock layer flow field and on the centerbody surface. In the air test slight flow separation occurs upstream of the leading edge of the fins and a secondary bow shock forms around each projecting fin. This shock is weak and not well defined close to the centerbody surface. However, it becomes quite strong about the portions of the fin that extend further out into the shock layer. A strong vortex flow occurs at the fin-centerbody stagnation junction. This is due in part to the pressure gradient along the fin stagnation line and it produces here a downward flow towards the centerbody surface. This reverse flow towards the centerbody occurs along the fin stagnation line and to a lesser extent along the fin surface. The resulting stagnation region vortices and secondary bow shock influence the primary centerbody shock layer flow and cause, 1, divergence of the primary boundary layer flow and, 2, surface pressure increases at considerable distances from the fin-centerbody junction.

Pressure measurements taken in the shock tunnel, again at M_∞ = 12, on a centerbody with rod-type (hemicylindrical leading edge, flat plate) control also indicate in this case that combined flow separation, vortex and secondary bow shock wave phenomena are present. That is, an increase in centerbody surface pressure of approximately a factor of two is experienced at fair distances from the rod-centerbody junction. The integrated effect of this centerbody surface pressure increase as well as the induced pressures on the rod leading edge produce a control force.

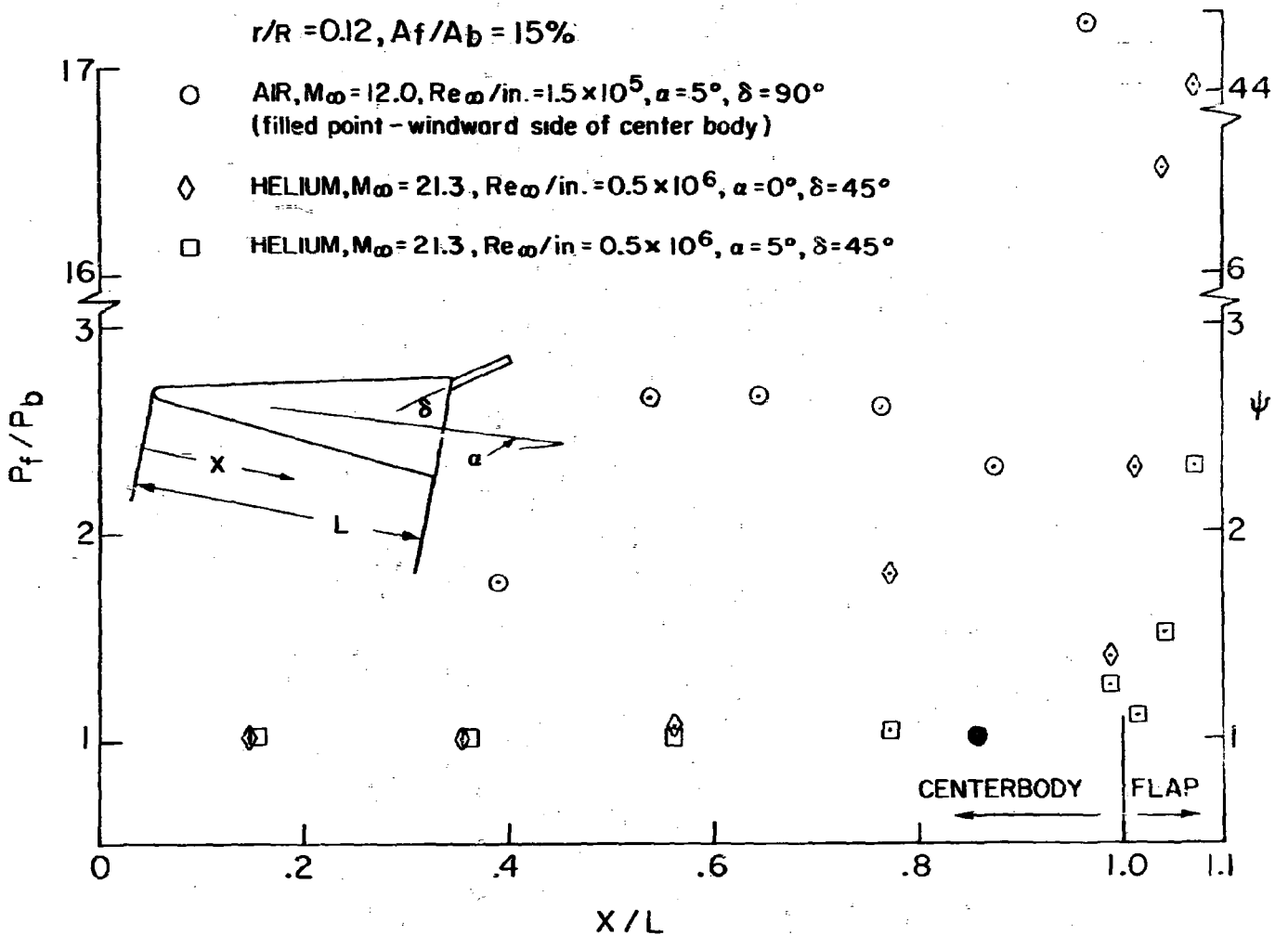


Fig. 10. Flap Influence on Centerbody Surface Pressure Distribution, 9° Sphere Cone

Results from flow visualization experiments conducted in helium at $M_\infty = 21.3$ and Re_∞ /in. $= 0.5 \times 10^6$ are presented in Fig. 11. Shown are a series of schlieren photographs which indicate the shock layer flow field geometry for a blunt model (a 9° sphere-cone with a nose bluntness ratio of .12) with a base-mounted flap ($\delta = 45°$) at various centerbody angles of attack (0° to 12.5°). Oil streak photographs for the same test series are also shown in the figure for the plane view of the fin side (suction side) and the side view of the model. At $\alpha = 0°$ the centerbody shock layer impinges on the flap and a strong reverse flow exists on the centerbody in the vicinity of the flap and for a considerable distance further forward on the centerbody (35). This is compatible with the aforementioned pressure data. As the centerbody angle of attack increases the reverse flow subsides. However, boundary layer separation occurs due to the presence of the flap.

These separation phenomena have been observed in more detail on two-dimensional sharp leading edge models where the wedge angle (θ) and the flap angle (β) were both varied over a wide range of values. In these tests, pressures were measured on the model surface ahead of and on the flaps. In addition, oil streak experiments were conducted and schlieren documentation of shock wave-boundary layer relationships was obtained. Choosing from these experiments the combination of θ and β values at which separation occurs at a fixed distance from the leading edge (1/2"), the curve of Fig. 12 may be drawn. It can be seen that the thick boundary layers are more conducive to flow separation at small values of θ. The reduction in θ values is equivalent to increasing the centerbody angle of attack (α). Carrying this analysis to the axisymmetric case, it is concluded for a flap-controlled vehicle at angle of attack and having a thick, laminar boundary layer that large flap deflections (β or δ) will cause increased separation of the boundary layer. This would imply under viscous flight conditions that the flap would have limited effectiveness as an attitude control device. The extent to which this is true is discussed later.

From the above experiment a model of the flow mechanism over the wedge with flap was derived and is shown sketched in Fig. 13. The separated region may be divided into two zones: 1, a quiescent zone immediately after the separation; and, 2, a turbulent zone ahead of the reattachment. The boundary layer was thinner in the latter zone and a recirculation flow existed and is shown both in top and side views (Fig. 13). The severe non-two-dimensionality of the flow over the wedge and particularly of the flow reattachment on the flap is also shown.

Pressures measured on these models show, as might have been expected from other work, a pressure plateau over the separated region. The magnitude of the pressure plateau corresponds very well to inviscid wedge pressures computed from the measured angle at the edge of the separated

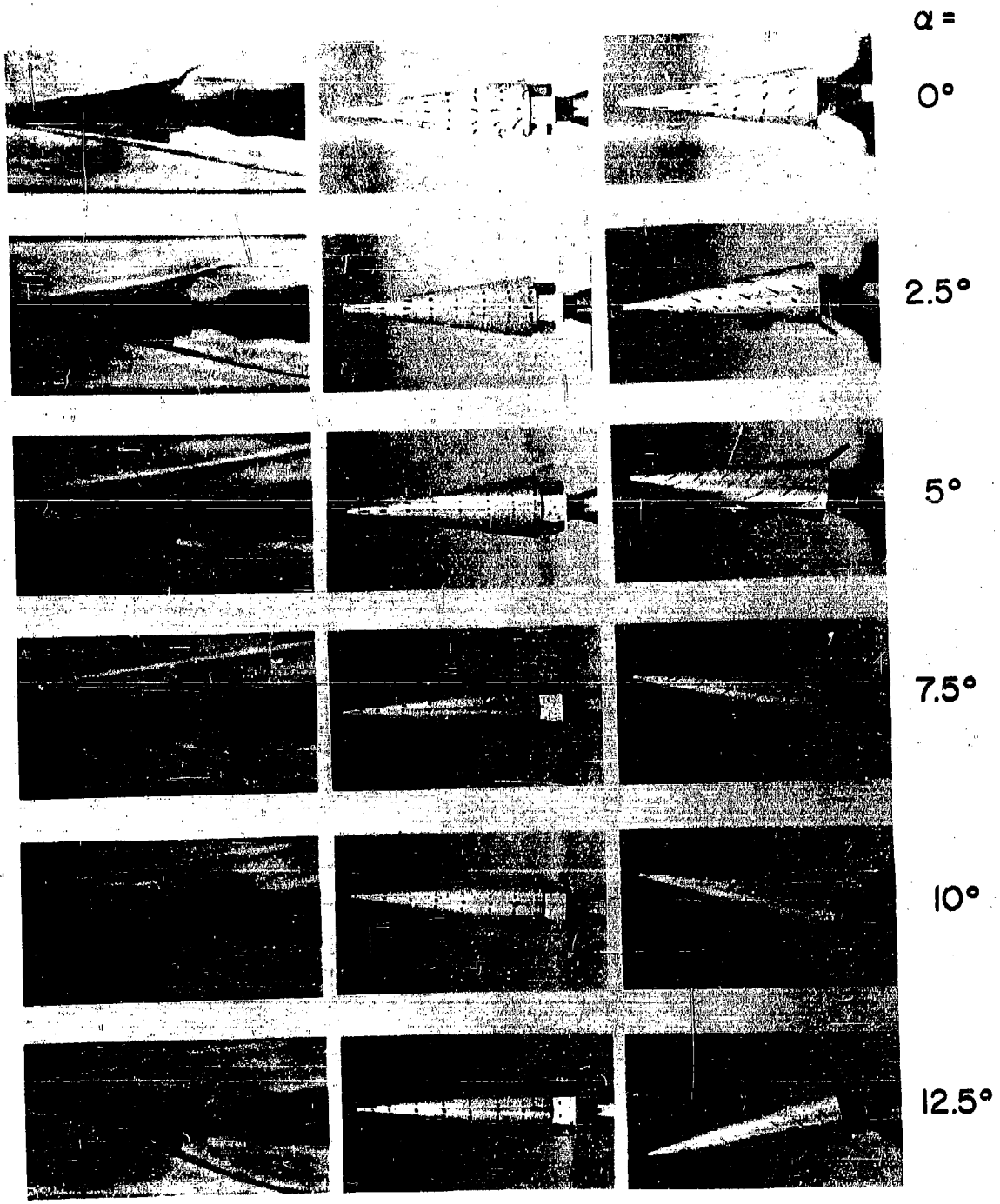


Fig. 11. Schlieren and Oil Flow Data, 9° Sphere-Cone with Flap

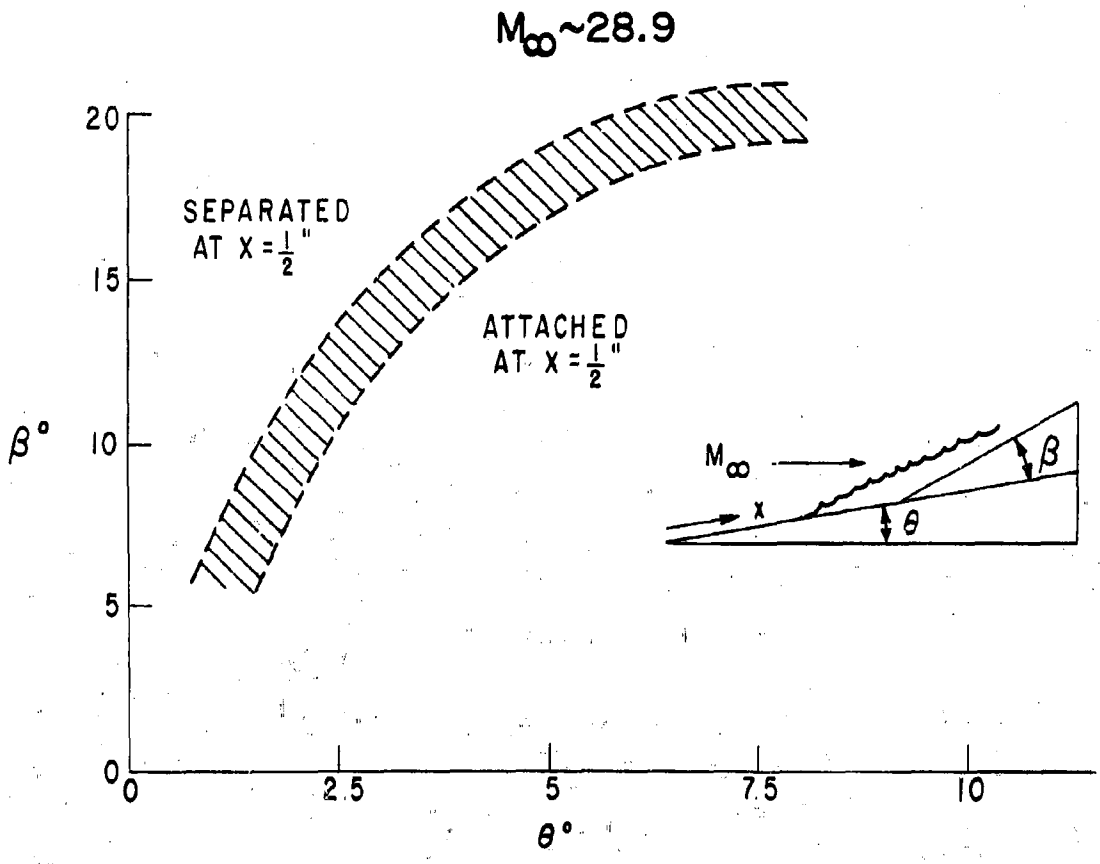


Fig. 12. Wedge-Flap Boundary Layer Separation (Helium)

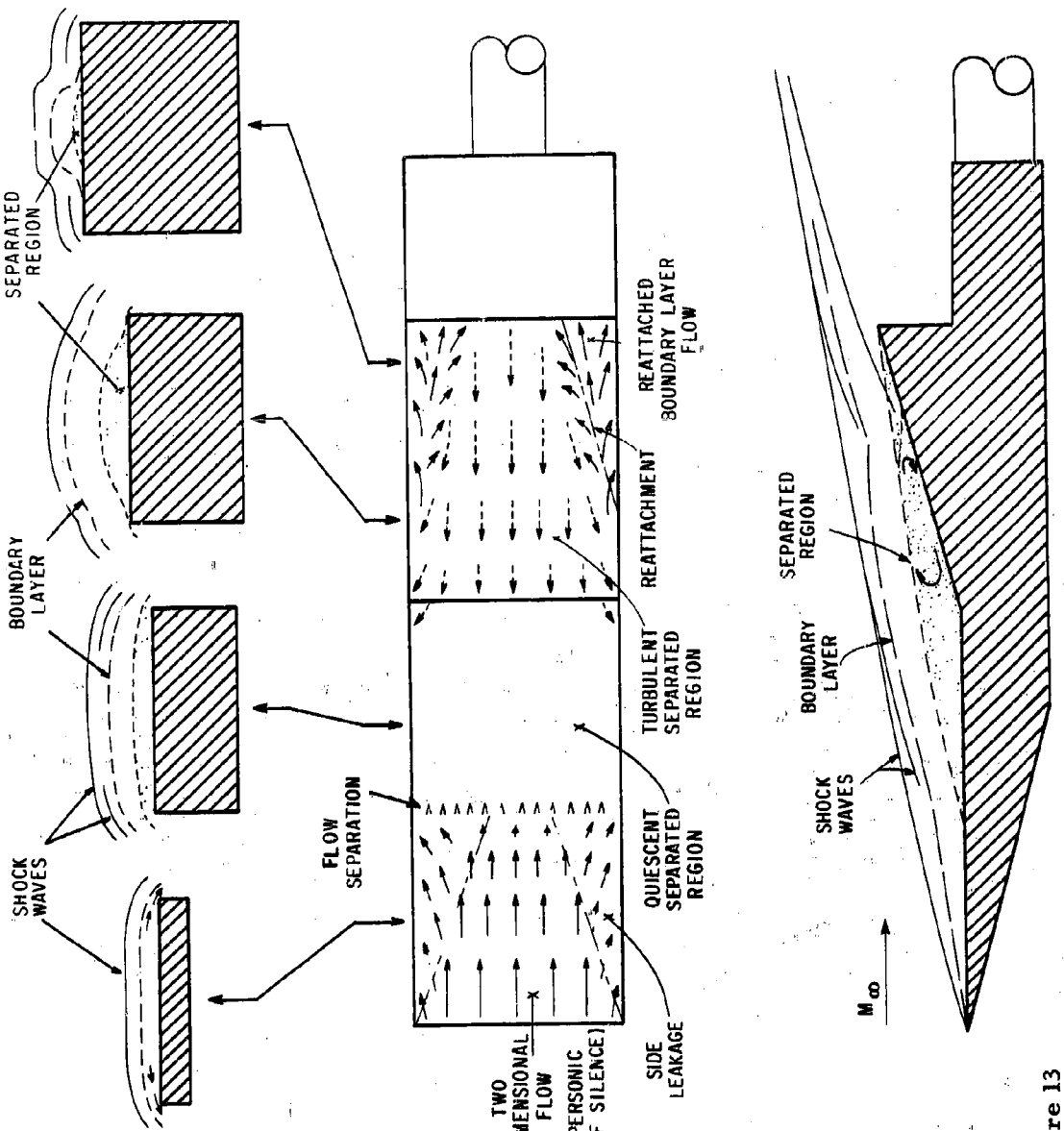

SEPARATED REGION
BOUNDARY LAYER
SHOCK WAVES
FLOW SEPARATION
REATTACHED BOUNDARY LAYER FLOW
REATTACHMENT
TURBULENT SEPARATED REGION
QUIESCENT SEPARATED REGION
SIDE LEAKAGE
TWO DIMENSIONAL FLOW
HYPERSONIC (ZONE OF SILENCE)
SEPARATED REGION
BOUNDARY LAYER
SHOCK WAVES
M∞

Figure 13

boundary layer. The flap pressures show a very strong dependence on wedge angle.

The foregoing data were obtained as stated on a sharp leading edge wedge. It is to be remembered that the phenomena which occurred have a strong dependence on leading edge bluntness (36). In a like manner the extrapolation of these data to an axisymmetric body requires among other considerations the consideration of nose bluntness. However, the pressure, oil flow and schlieren data of Figs. 10 and 11 exhibit the same qualitative trends as the two-dimensional data. At small α values (large θ values) the thick boundary layer flow separates on the centerbody and then reattaches on the flap surface. The shock layer also impinges on the flap producing a shock and relatively high surface pressures near the tip of the flap. The centerbody and flap separated region is turbulent with a reverse flow which causes relatively high pressures upstream on the centerbody. As the angle of attack increases (θ decreases) a quiescent separated region occurs on the centerbody in front of the flap and the pressure on the flap is small. There is no evidence of the oil flowing in this region (Fig. 11). The boundary layer may reattach near the tip of the flap, but the inviscid shock layer does not impinge upon the flap. Strong three-dimensional effects exist back on the flap due to the cross-flow around the axisymmetric body at angle of attack and due to edge effects on the two-dimensional flap such as observed in the wedge-flap tests (Fig. 13).

The above flow separation observations being made in helium are more severe than would be expected for the same conditions in air. An approach to determining this difference is given in the appendix (also see ref. 24).

The effectiveness of a flap in achieving attitude control was further investigated by measuring the static stability characteristics for a centerbody with low zero lift drag characteristics and at more viscous free-stream conditions than in the previous 0.3 bluntness, 9° sphere-cone tests. The centerbody was chosen from the family of shapes shown in Fig. 4. In Fig. 14 are presented the zero angle of attack drag data obtained for two of the three basic shapes as a function of nose bluntness. Also included in this figure is the inviscid exact flow field prediction for a 9° sphere-cone. The data show that strong viscous effects are encountered at the test condition under consideration. The data further show that a decrease in drag is obtained due to slight blunting of the model nose. The cone data obtained in this study are presented again in Fig. 15 along with other cone data (25) (26). The data are correlated in terms of $C_{D_V} M_\infty^2$ (viscous drag) and the interaction parameter $M_\infty^3 \sqrt{C} / Re_{x_L}$ (i.e. $\bar{\chi}$). The general trend is for the sharp cone body to yield higher C_{D_V} values than the blunter bodies. The data from the shock tunnel study (Fig. 14) when correlated in this fashion show that the viscous drag decreases with increasing bluntness over the range from $r/R = 0$ to

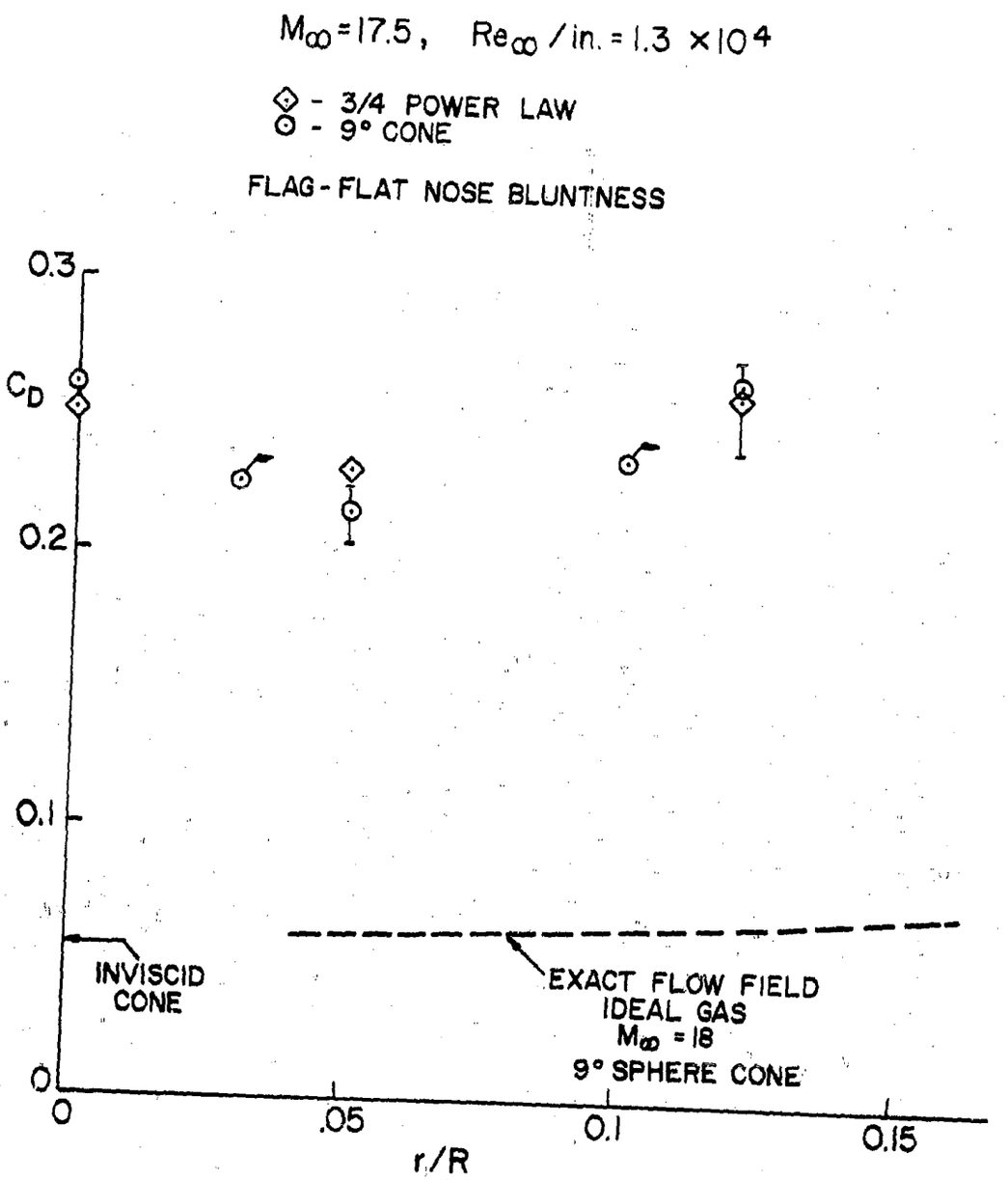


Fig. 14. Zero Lift Drag vs. Nose Bluntness (Air)

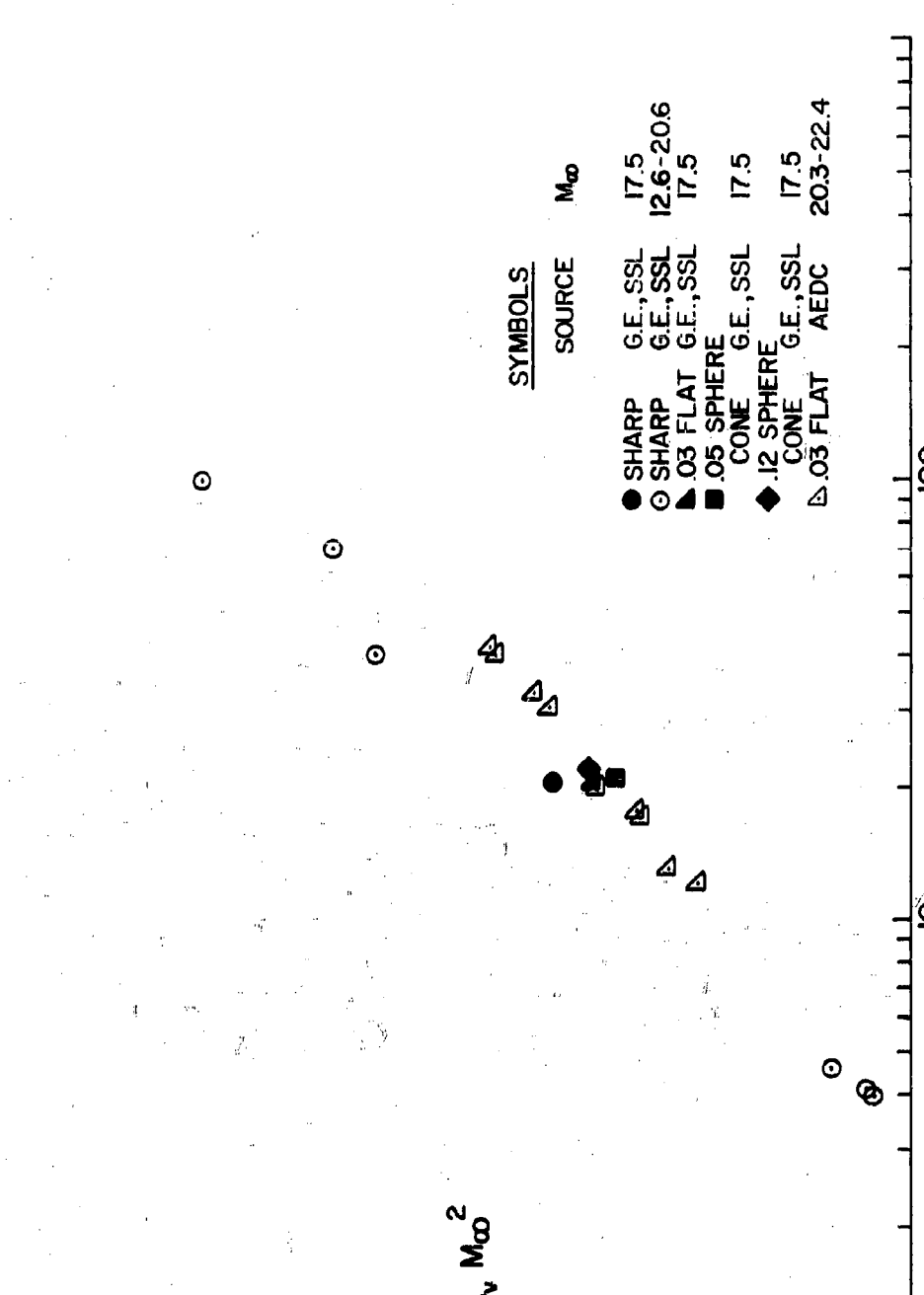


Fig. 15. Zero Lift Viscous Drag, 9° Cone and Sphere-Cone (Air)

$^{r}/R$ =.05. Fig. 14 also shows that the total zero lift drag for both the sphere-cone and 3/4 power law body is a minimum at the $^{r}/R$ = 0.05 value.

The $^{r}/R$ = 0.05 bluntness cone, tangent-ogive, and 3/4 power law bodies were also tested at M_{∞} = 17.5 over an angle of attack range from 0° to 13°. The static stability aerodynamic coefficients, $^{L}/D$ and Xc.p./ℓ obtained for the sphere-cone and 3/4 power law bodies are presented in Fig. 16. Both the sphere-cone and 3/4 power law bodies give similar $^{L}/D$ values in this α range. The 3/4 power law body $^{L}/D$ is somewhat lower due primarily to its lower C_L values. The C_D values for the two bodies are approximately the same over the given α range. Again the approximate correction was applied to the Newtonian $^{L}/D$ prediction for this body (27). Its agreement with the experimental data is as for the higher Reynolds number test condition (Fig. 8). The two bodies have statically unstable center of pressure characteristics. In the earlier study of a finned body (1) it was shown that with initially unstable bare body static stability characteristics desirable attitude control characteristics could be obtained with a body after a control surface was placed on its aft portion. Also shown in Fig. 16 are the static stability characteristics for a 0.05 nose bluntness, 3/4 power law body with a single flap located on the low pressure side of the body. The addition of a flap results in a small reduction in $^{L}/D$. This reduction diminishes with increasing centerbody angle of attack. However, the flap does produce the desirable statically stable center of pressure characteristics.

These flap control results are in agreement with the results from the earlier boundary separation study; that is, at the viscous flight conditions as the angle of attack increases the thick boundary layer separates and reduces the effectiveness of the flap. This is principally due to the fact that only very small induced surface pressures are then produced. These induced pressures are concentrated on the flap and near the end of the centerbody, thus accounting for the observed Xc.p./ℓ vs α characteristics. At the less viscous test condition (Fig. 8) boundary layer separation was slight, strong shocks were formed, high pressures occurred on the control surface, control surface-centerbody interactions resulted in relatively high induced centerbody surface pressures over a large area well upstream of the aft end of the body and effective attitude control was thus achieved.

Magnetohydrodynamic Control

Presented here are some of the experimental results from the feasibility study of the MHD control concept.

Theoretical analyses pertinent to the two-dimensional MHD channel problem have been previously performed (28, 29, 30). Thus, only a brief presentation of the basic equations pertinent to this problem will be included

$M_\infty = 17.5, Re_\infty = 1.3 \times 10^4/in.$

□ 3/4 POWER LAW BODY, r/R = .05

○ 9° SPHERE CONE BODY, r/R = .05

◇ 3/4 POWER LAW BODY, (r/R = .05) WITH FLAP (δ = 30°)

\+ NEWTONIAN WITH APPROX. CORRECTION

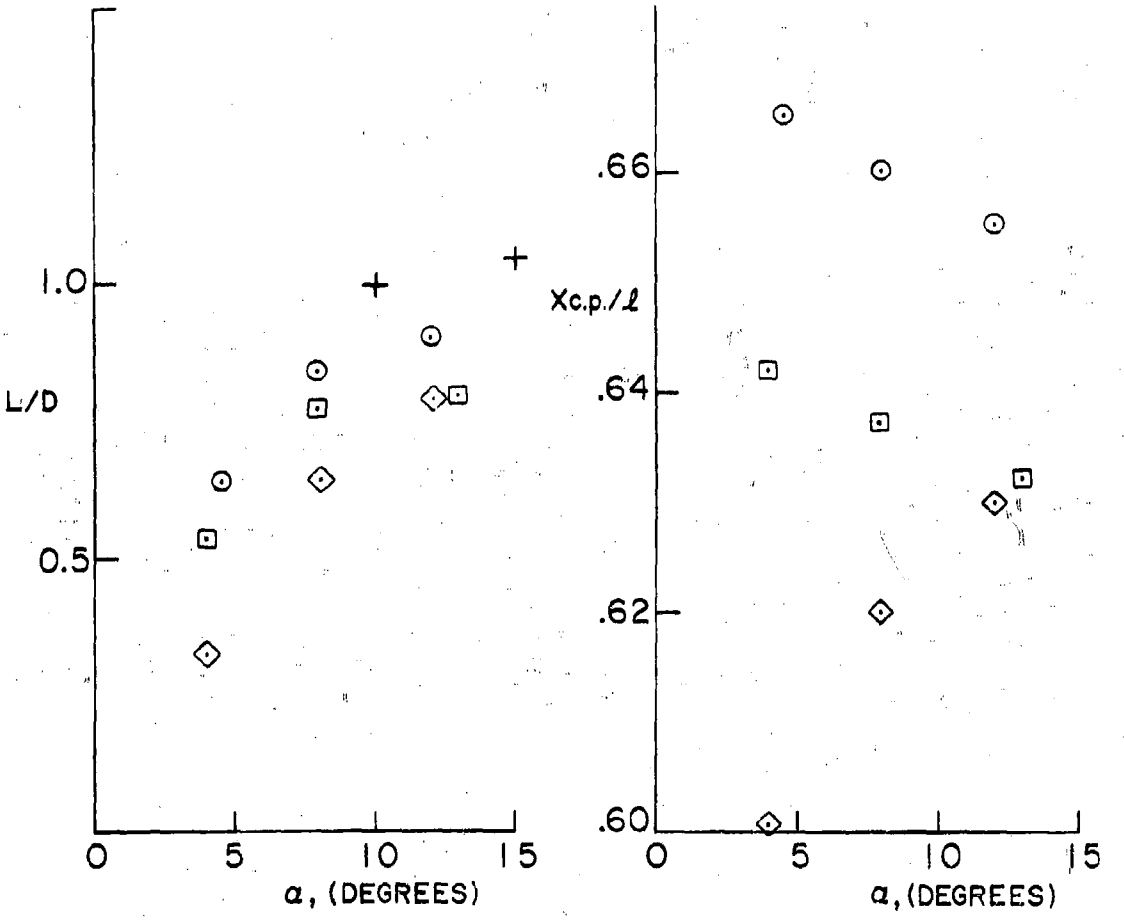


Fig. 16. Static Stability Data (Air)

here. These equations point out the important plasma properties of interest and give some insight into the experimental design problems encountered in the experiment. The gas dynamic equations are:

(conservation of momentum)

$$\rho u \frac{du}{dx} = -\frac{dp}{dx} + \sigma (E - uB) B \tag{1.0}$$

(conservation of mass)

$$\rho u h = \text{constant} \tag{1.1}$$

(conservation of energy)

$$\rho u c_p \frac{dT}{dx} + u \frac{dp}{dx} + \sigma (E - uB)^2 \tag{1.2}$$

and (perfect gas equation of state)

$$p = \rho R T \tag{1.3}$$

An additional equation accounts for the fact that an essentially non-conducting plasma is flowing above the channel but adjacent to the conducting plasma in the channel. This is due to the fact that the flow in the channel is seeded to increase its electrical conductivity. The relationship between the pressures in the two plasma layers across a dividing streamline is:

$$p - p_1 = \frac{\rho_1 u_1^2}{\beta} \frac{dh}{dx} \quad \text{where } \beta = \sqrt{M_1^2 - 1} \tag{1.4}$$

where in the preceding equations h is the height or thickness of the electrically conducting plasma layer and the M_1, P_1, u_1, and ρ_1 quantities apply to the non-conducting plasma.

The magnetic and electrical parameters when applied yield the following relationships of interest:

(Ohm's Law)

$$\bar{J} = \sigma \left\{ \frac{\bar{E} + \bar{E}'' + \bar{v} \times \bar{B}}{1 + \omega^2 \tau^2} - \frac{e\tau}{B(1+\omega^2\tau^2)} (\bar{E} + \bar{E}'' + \bar{v} \times \bar{B}) \times \bar{B} \right\} \tag{2.0}$$

when the conducting plasma is fully ionized gas. When the plasma is partially ionized this equation becomes:

$$\bar{J} = \frac{1}{K}\left\{ \sigma(\bar{E} + [1-fb]\bar{E}'') - \frac{f^2}{K_- + K_+}\frac{\sigma}{B}\bar{E}'' \times \bar{B} \right.$$
$$\left. - \frac{\omega\tau}{BA}(1-2fb)\left[\sigma\bar{E}' \times \bar{B} + \sigma(1-fb)\bar{E}'' + \frac{f^2}{K_- + K_+}\sigma B\bar{E}''\right]\right\} \quad (2.1)$$

The constants K, K_-, K_+, A, f, and b are defined based upon electron-ion, electron-neutral and ion-neutral collisions, and both electron and ion spiraling. In the actual experimental study the flow is partially ionized. However, the experiment will be discussed here only in general and therefore in terms of the simpler equation, that for the fully ionized gas.

In the experimental set-up an electromagnet is rigidly mounted in the blunted wedge model shown in Fig. 5. Also shown are several of the electrodes used in this study. The electrodes were 4" long and 1/2" high. They were mounted parallel and 3" apart in the channel, held in place by tungsten clamps and protected from severe mechanical loading by covers. Provisions were made for externally heating these electrodes with a 3 KW transformer. In this manner electrode surface temperatures as high as 1300°C were obtained. The centerbody is a blunt leading edge wedge with a 17° half angle and a leading edge radius to base height ratio of 0.0375. The magnet is energized by an external power supply capable of putting out a 1.5 megawatt pulse, resulting in a 10 ms duration, 14,000 gauss field in the electromagnet. This field is mutually perpendicular to the direction of plasma flow and the model wall or channel floor. The resulting induced electromotive force is illustrated as shown in the insert of Fig. 5 for the case of the magnetic field directed into the surface of the page or model. In the experimental study the composite model with magnet and electrode components is mounted on a drag balance which is attached to the sting support system located in the 30" shock tunnel test section. When flow was established over the model a seed material was introduced into the model's stagnation region to increase the electrical conductivity level of the shock layer which passes through the channel. The seed material in this experiment was potassium bicarbonate which has a low ionization potential. Air was used throughout as the test gas. The potassium bicarbonate was formed into a paste (by using acetone) and coated into a relatively small rectangular slot located on the model surface parallel to and just downstream of the stagnation point. The seed material aerodynamically heated in this stagnation zone ionizes, and then is swept downstream through the two-dimensional channel. A boundary layer trip was provided downstream of the slot in order that all of the seed material would not be trapped in the boundary layer and, therefore, the plasma flowing in the trough would be seeded to a depth sufficient to

completely immerse the electrodes. Measurements taken downstream of the slot-boundary layer trip arrangement with a plane grating f/6.5 spectrograph showed that adequate depth of seeding could be accomplished with this technique.

The tests were conducted at conditions that had the following calculated properties: $T_s = 5300^{\circ}K$, n_{e_s} (unseeded) = 10^{14}/cc and $p_s = 1.2$ atm. The calculated plasma properties of interest based on equilibrium flow field computations (31) in the two-dimensional trough were $p_c = 0.1$ atm., $T_c = 2500^{\circ}K$, $\rho_c = 6 \times 10^{-6}$ gm/cm^3, and $u_c = 3 \times 10^5$ cm/sec. The test section free stream flow Mach number was 4.7. The normal duration of test flow is 1.0 millisecond (8).

In addition to attempting to obtain aerodynamic data, efforts were made to make electrical measurements as well. The prime scalar electrical quantity measured in this experiment was:

$$J_y = \frac{\sigma}{1+\omega^2\gamma^2}\left(\omega\gamma E_x'' + E_y'' - \gamma_x B_z\right) \qquad (2.2)$$

This current crossed with the applied magnetic field yields the Lorentz force used to develop the required attitude control, i.e.,

$$\bar{F} = \bar{J} \times \bar{B} \qquad (3.0)$$

It can be seen from the equations that optimum performance may be achieved if $\omega\gamma$, E_x'' and E_y'' are small and σ, u_x and B_z are large. Of course limiting values must be placed on B_z since the ω term is directly proportional to the magnetic field strength. In the experiment to be discussed the initial conditions are $\bar{E} = 0$, $\bar{B} = \bar{k}\, B_z$, $\bar{u} = \bar{\imath}\, u_x$ and $\bar{E}'' = \bar{\imath}\, E_x'' + \bar{j}\, E_y''$. The magnitude of the current ($\bar{J}$) may be further controlled by the application of a variable external resistive load (R_L) in series with the electrodes. When $R_L = 0$, max. current and force are obtained. When R_L equals the MHD channel's internal electrical impedance, max. electrical power may be extracted. The load resistors used in this experiment were special IRC* low inductance, pulse-viewing resistors.

Assuming that the ionized species produced in the seeded plasma at the model stagnation point would undergo a frozen expansion in flowing around the model and through the channel, the calculated electrical conduc-

* International Resistor Corp., Phila., Pa.

tivity of the plasma would then be 300 mhos/m. With B = 10,000 gauss, the maximum power density would be 675 watts/cm^3, * and the maximum Lorentz force would be approximately 20 lbs, a respectable force in this experiment for achieving attitude control. The actual experimental results are given in Fig. 17. Two sets of data from this experiment are presented. With both electrodes at room temperature and B = 10,000 gauss the measured peak power density is 1.6 W/cm^3 and the electrical conductivity value is 2.9 mhos/meter (32, 33). Heating the positive electrode (this electrode was La B_6 fused on a rhenium strip) to 1300°C resulted in an increase in channel electrical conductivity to approximately 20 mhos/meter and an increase in the peak electrical power density to 11.4 watts/cm^3. Assuming $u_c = 3 \times 10^5$ cm/sec the resulting Lorentz force would then be approximately 1.3 lbs; this is below the reading resolution of the piezoelectric force balance used in the experiment (the aerodynamic force is approximately 100 lb). The same problem occurred for the piezoelectric quartz pressure transducers used in monitoring the channel pressure. Thus the measured force and pressure data showed no resolvable effects due to the MHD phenomena; this is compatible with the measured power densities.

It is important to note again that calculations based on an equilibrium expansion of the flow from the stagnation line through the channel yield a channel plasma temperature of approximately 2500°K and a resulting $\sigma \sim$ 25 mhos/meter for the seeded flow. The latter value is in reasonable agreement with the generator electrical conductivity calculated from the experimental data when the positive electrode was externally heated to 1300°C.

CONCLUSIONS

An experimental evaluation of flap, fin, rod and MHD control concepts was made. Data were obtained in air and helium covering a free-stream Mach number range of 4.7 to 28.9 and a free-stream Reynolds number per inch range of approximately 10^4 to 10^6.

The presence of the control surfaces -- that is, the flap, fins and rod -- on the axisymmetric, slender, blunt centerbodies caused flow interactions in the centerbody shock layer which could result in the formation of strong and weak shocks, vortices, flow separation and an increase in pressure over an extensive area of the centerbody surface. Without the control surfaces present the measured surface pressure distribution on the 9° sphere-cone centerbody in both air and helium was in very good agreement with exact flow field predictions. The control surfaces, however, did produce increases in centerbody surface pressure of as high as a factor of 17 above the body alone pressure value for the higher Reynolds number test condition

* This supersedes the erroneous value given in reference 32.

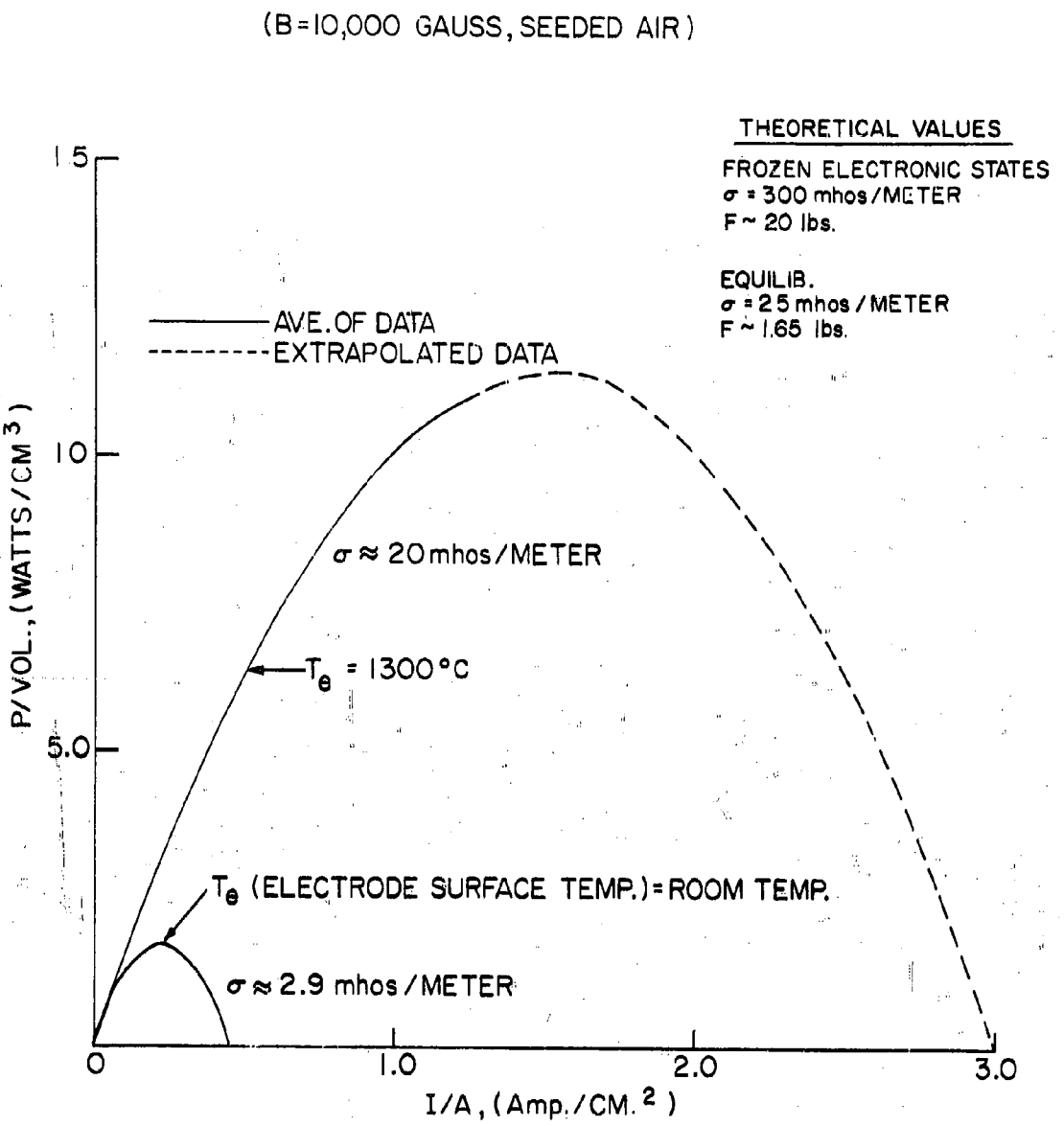


Fig. 17. MHD Attitude Control - Two-Dimensional Channel Exper. Electrical Performance

in air. The centerbody pressure increase is found to be due partially to boundary layer separation but mainly to the formation of shocks and vortices in the centerbody shock layer. These control surface-centerbody flow interactions in the high Reynolds number case combined to produce an integrated force useful in generating flight attitude control. In the lower Reynolds number helium tests moderate pressures were measured on the flap control surface. The helium data further served to indicate the occurrence of boundary layer separation and reattachment as well as impingement of the shock layer on the flap. In the helium study where a relatively thick laminar boundary layer was present, the increase in centerbody surface pressure at $\alpha = 0^{\circ}$ was due mainly to turbulent feedback into the separated region caused by the larger pressures on the flap. As α increases separation occurs further upstream on the centerbody. The boundary layer either does not reattach on the flap or reattaches without the centerbody shock layer impinging on the flap. This leads to a reduction in control effectiveness.

The comparative study of flap, fin and rod static stability characteristics showed that appreciable control effectiveness might be expected under statically stable trim conditions with flap or rod attitude control for the higher Reynolds number condition. An unswept leading edge fin system produced statically unstable and undesirable flight characteristics at $\alpha > 8^{\circ}$. However, an earlier shock tunnel study determined how satisfactory performance could be obtained with this system by merely varying the fin design parameters (such as sweeping the fin leading edge).

For the bare 3/4 power law and 9° cone bodies it was found that small blunting of the nose reduced the total zero lift drag presumably by reducing the viscous drag contribution. The viscous drag contribution when used as an approximate correction term in the bare body Newtonian L/D prediction yields an improved agreement between experiment and theory. The phenomenological analysis of the schlieren pressure and oil flow data in helium supports the low Reynolds number, flap control, static stability data obtained in air.

The experimental study in the MHD area resulted in the design and successful testing of a two-dimensional MHD channel located on a blunt wedge model. The limited aerodynamic and electrical measurements obtained indicate that, under the given test conditions, small induced forces were obtained but that the performance of the channel could be predicted using basic MHD equations assuming equilibrium flow around the model. An encouraging result was that the channel was a fairly good electrical power generator yielding a peak power density of 11.4 watts/cm^3 when one electrode was heated to 1300°C.

APPENDIX

HELIUM-AIR SIMULATION OF HYPERSONIC BOUNDARY LAYERS AND FLOW SEPARATION

Following the treatment of Ref. (34) one may express the displacement thickness for a flat plate under adiabatic conditions with Prandtl number = 1 and $M_\infty >>> 1$ as

$$\frac{\delta^*}{L} \sim (\gamma-1)\frac{M_\infty^2}{R^N}$$

where L is distance from leading edge and N is the incompressible skin friction law exponent.

If we wish to simulate Mach number and Reynolds number between helium and air we must have

$$\frac{\delta^*_{He}}{\delta^*_{AIR}} \sim \frac{(5/3-1)}{(7/5-1)} = \frac{5}{3}$$

From this we see that since processes such as flow separation are more prone to occur in thicker boundary layers with all other flow properties equal (M, Re,) the flow separation measured in the present helium experiments are more severe than would be the case with air as the flow medium. As a first approximation, one may say the separation point S measured from the leading edge is inversely proportional to δ^*. So,

$$\frac{S_{He}}{S_{AIR}} \sim \frac{\delta^*_{AIR}}{\delta^*_{He}} \sim \frac{3}{5}$$

Consideration of the ratio of model wall temperature to test section free-stream temperature in the helium tunnel and shock tunnel test results in even a bigger value for the simulation. This simulation for boundary layer thickness may be expressed as

$$\frac{\delta_{He}}{\delta_{AIR}} \sim \frac{(\gamma-1)_{He}}{(\gamma-1)_{AIR}} \frac{(T_{ad.}/T_e)_{He}}{(T_e-T_w/T_e)_{AIR}}$$

where $T_{ad.}$ is the adiabatic model wall temperature in the helium test

T_e is the respective temperature at the edge of the boundary layer for helium or air

T_w is the ambient model wall temperature in the shock tunnel air test.

SYMBOLS

A	--	area
$\overline{B}$	--	magnetic field strength and direction
$\sqrt{C}$	--	$\mu_w T_\infty / \mu_\infty T_w$
C_{Dv}	--	aerodynamic viscous drag
D	--	aerodynamic drag
D_B	--	base diameter
$\overline{F}$	--	Lorentz force
$\overline{J}$	--	induced electric current
K	--	nose drag
L	--	aerodynamic lift, also model length
M	--	Mach number
N	--	Incompressible skin friction law exponent
P	--	pressure
R	--	base radius
R_e	--	Reynolds number
R_L	--	external load resistance
X	--	distance along centerbody axis
$X_{c.g.}$	--	center of gravity
$X_{c.p.}$	--	aerodynamic center of pressure
d_n	--	nose diameter
ℓ	--	model length
n_e	--	electron density
r	--	nose radius
$\overline{u}, \overline{v}$	--	shock layer flow velocity
α	--	centerbody angle of attack
$\delta(\beta)$	--	control surface perturbation parameter 1. deflection angle with respect to body axis (flap & fin) 2. distance into the centerbody shock layer (rod)
ϵ	--	$(\gamma-1)/(\gamma+1)$
χ	--	$M_\infty^3 \sqrt{C} / \sqrt{Re_{\infty_L}}$
$\omega\tau$	--	Hall parameter (MHD)
ψ	--	ratio of the measured flap pressure to the pressure measured at the most rearward centerbody station.

Subscripts

b	--	bare body (also base of centerbody)
c	--	MHD channel conditions
f	--	with flap
s	--	stagnation point
∞	--	free stream

REFERENCES

1. Harris, C. J. and Warren, W. R., "An Experimental Study of the Aerodynamic Characteristics of Finned Re-entry Vehicles," Proc. of the 6th Symposium on Ballistic Missiles and Space Tech., August, 1961, Also GE, TIS R61SD173, November 16, 1961.

2. Warren, W. R., Kaegi, E. M. and Geiger, R. E., "Shock Tunnel Experimental Techniques for Force and Moment and Surface Flow Direction Measurements," J. Am. Rocket Soc., January 1961.

3. Geiger, R. E., "Experimental Lift and Drag of a Series of Glide Configuration at Mach Number of 12.6 and 17.5," Jour. of the Aerospace Sci., Vol. 29, No. 4, April 1962

4. Warren, W. R., Kaegi, E. M., and Harris, C. J., "Results of Experimental Heat Transfer and Pressure Distribution Studies on Re-entry Vehicle Control Surfaces," Proc. of the 4th Symposium on Space and Ballistic Missile Tech., 1959 also GE TIS 59SD754, August, 1959.

5. Johnson, R. H., "Effects from Thin Fins in Thick Hypersonic Boundary Layers," Am. Roc. Soc. Meeting. Preprint 2096-61, At 1961 and GERL 61-RL-29036, 1962.

6. Fitzgibbon, S. A., "Flow Field Characteristics of Low-Drag Bodies of Revolution at Mach 18," GE MSVD, ADM 1:70, March 1962.

7. Harris, C. J., Kaegi, E. M. and Warren, W. R., "Pressure and Force Transducers for Shock Tunnels," ISA Jour., August 1960; also GE TIS R59SD327, Jan., 1959.

8. Harris, C. J., "Some Aspects of a Shock Tunnel Study Involving the Application of MHD Forces to Vehicle Attitude Control," GE MSVD Adv. Aero. Lab. TM 98, Feb. 6, 1960.

9. Warren, W. R., "The Design and Performance of the General Electric Six-Inch Shock Tunnel Facility," Proc. of 1st Shock Tube Symposium, USAF Special Weapons Center, SWR-TM-57-2, Feb., 1957.

10. Kaegi, E. M., Warren, W. R., Harris, C. J., and Geiger, R. E., "The Capabilities of the Shock Tunnel in the Study of the Aerodynamics of Atmospheric Entry," G.E. R62SD56, May, 1962.

11. Johnson, R. H., "Some Influence of Fluid Viscosity on High Mach Number, Low-Density Wind Tunnel Operation," GE Report 61-RL-2610C, January, 1961.

12. Tsien, Hsu-shen, "Similarity Laws of Hypersonic Flows," Jour. of Math and Physics, Vol. 25, No. 3, Oct., 1946.

13. Probstein, R. R., "Interacting Hypersonic Laminar Boundary Layer Flow Over a Cone," TR AF 2798/1 Div. of Eng., Brown Univ., March, 1955.

14. Cheng, Hsien K., "Similitude of Hypersonic Real Gas Flows Over Slender Bodies with Blunted Noses," J. Aero. Sci., Vol. 26, No. 9, September, 1959.

15. McCauley, W. D., "Low Density Hypersonic Similarity of Cones and Slightly Blunted Sphere Cones," GE TIS 62SD540, May, 1962.

16. Geiger, R. E. and Harris, C. J., "Experimental Investigation of the Hypersonic Characteristics of a Series of Basic Lifting Shapes," GE TIS R62SD3, January 1962.

17. Cheng, H. K., "Hypersonic Flow with Combined Leading-Edge Bluntness and Boundary-Layer Displacement Effect," Cornell Aero. Lab., Report AF-1285--A-4, August 1960.

18. Gravalos, F. G., Edelfelt, I. H., and Emmons, H. W., "The Supersonic Flow About a Blunt Body of Revolution for Gases at Chemical Equilibrium," Proc. of 9th Internat. Astronautical Congress, Amsterdam, 1958, also GE TIS R58SD 225, June 1958.

19. Stamm, S. L., "Flow Field Data on Several Sphere-Cone Configurations" PIR 538-888-, GE MSVD August 1959.

20. Johnson, R. H., and Bigelow, E. C., "Separation and Three-Dimensional Flows in Hypersonic Boundary Layers," GERL # 62-RL-3037C, May, 1962.

21. Cowling, T. G., "Magnetohydrodynamics," Interscience Publishers, Inc., 1957.

22. Sears, W. C., "Some Solutions of the Macroscopic Equations of Magnetohydrodynamics," ARS preprint, 879-59, August 1959.

23. Resler, E. L. and Sears, W. R., "The Prospects for Magneto-Aerodynamics," Jour. of Aero-Sciences, Vol. 25, No. 4, April 1958.

24. Henderson, A., "Experimental Investigation of the Air-Helium Simulation Problem on Various General and Specific Configurations at Hypersonic Speeds," Proc. of the 2nd National Symposium on Hypervelocity Techniques, March 1962.

25. Geiger, R. E., - G. E. Space Sciences Laboratory, private communication.

26. Lukasiewicz, J., Whitfield, J. D., and Jackson, R., "Aerodynamic Testing at Mach Numbers from 15 to 20," ARS preprint 1969-61, August, 1961.

27. Studerus, C., G. E. MSD private communication (provided inviscid Newtonian calculation.)

28. Sutton, G. W., "Design Considerations of a Steady DC MHD Electrical Power Generator," GE R59SD432, September, 1959.

29. Brunner, M. J., "The MHD Re-entry Vehicle Channel," GE R61SD31, April 1961.

30. Harris, L. P. and Moore, G. E., "Some Electrical Measurements on MHD Channels," GE 62-RL-2961G/C, March, 1962.

31. Ridyard, H. W., "Comparison of the Ionized Shock Layer About Two and Three Dimensional Blunt Shapes at Hypersonic Speeds," GE MSVD TIS R60SD327, Feb., 1960.

32. Muntz, E. P., Harris, C. J., and Kaegi, E. M., "Techniques for the Experimental Investigation of the Properties of Electrically Conducting Hypersonic Flow Fields," Proc. of the 2nd National Symposium on Hypervelocity Techniques, Denver, Colo., March 1962.

33. Harris, C. J., "Initial Experimental Results of a MHD Flight Control Study, GE MSVD, ADv. Aero. TM 116, January 1961.

34. Love, E. S. and Henderson, A., "Some Aspects of Air-Helium Simulation and Hypersonic Approximation," NASA TN-D-49, October 1959.

35. Kaplan, M., "High Speed Boundary Layer Separation in a Compression Corner," GE TFM-HTT-8151-001, May, 1962.

36. Potter, J. L. and Whitfield, J. D., "Effects of Slight Nose Bluntness and Roughness on Boundary-Layer Transition in Supersonic Flows," Jour. of Fluid Mechanics, April, 1962.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the assistance of E. C. Bigelow in obtaining the helium data and the personnel under M. J. Engel who assisted in obtaining the air data. They would further like to thank Miss E. Miller and Mr. C. Smith for their efforts in reducing the air data.

TECHNIQUES FOR RADIATION MEASUREMENTS AND FLOW VISUALIZATION OF SELF-LUMINOUS HYPERSONIC WAKES

R. L. Taylor
J. C. Keck
W. K. Washburn
D. A. Leonard
B. W. Melcher, II
R. M. Carbone

Avco-Everett Research Laboratory
A Division of Avco Corporation
Everett, Massachusetts

ABSTRACT

Various techniques developed at the Avco-Everett Research Laboratory for observing the flow field behind hypervelocity bodies in ballistic ranges are discussed. These techniques fall into two categories: (1) photoelectric instruments for measuring the radiation intensity and its distribution in the wake, and (2) instruments for visualizing or photographing the flow field.

In the first category a photoelectric recorder is discussed. This instrument measures the radiation intensity as a function of distance downstream in the wake. Also, a wake scanner is described by which it is possible to repetitively scan the wake every few microseconds and obtain profiles of the radiation distribution in a plane nearly perpendicular to the flight direction.

In the second category drum camera techniques including a "race track" adaptation are described for photographing the self-luminous flow field. Data available from these drum camera techniques include gas particle velocity in the wake, shedding frequency of turbulent eddies, and width and rate of growth of the luminous wake. A schlieren technique is discussed in which the refractivity of the test gas is increased by using a small partial pressure of sodium vapor in the gas and taking the schlieren photograph with light of wavelength close to the sodium resonance line at 5896 Å. Good quality schlieren photographs have been obtained at densities less than 1% of atmospheric. The use of an image converter to take snapshot photographs of the self-luminous wake is discussed. Pictures showing the details of the luminous flow field in the base region have been obtained.

The description of these instruments is given, typical data are shown, and the types of information obtainable from the data are illustrated.

This work was supported jointly by Headquarters, Ballistic Systems Division, Air Force Systems Command, United States Air Force under Contract No. AF 04(694)-33 and the Advanced Research Projects Agency monitored by Army Ordnance Missile Command, United States Army under Contract No. DA-19-020-ORD-5476.

TECHNIQUES FOR RADIATION MEASUREMENTS AND FLOW VISUALIZATION OF SELF-LUMINOUS HYPERSONIC WAKES

R. L. Taylor, J. C. Keck, W. K. Washburn, D. A. Leonard
B. W. Melcher, II, and R. M. Carbone
Avco-Everett Research Laboratory
A Division of Avco Corporation
Everett, Massachusetts

INTRODUCTION

This paper describes several techniques developed at the Avco-Everett Research Laboratory (AERL) for investigating the wake generated by hypervelocity projectiles in a ballistic range. The importance of the study of the hypersonic wake rests in the fact that most of the energy dissipated by an object re-entering the atmosphere goes into the wake. The processes occurring during the redistribution of this energy into the random thermal motion of the air molecules produce the re-entry observables, such as optical emission and radar signature. The observable phenomena of the wake are a complex interaction of aerodynamic effects, chemistry, ionization, and radiation. Experimental study of these observables is needed in order to guide the development of analytical descriptions of the wake.

The ballistic range is the primary laboratory facility which has been used to simulate the wake phenomena of re-entry (1). Because of the low level of radiation intensity and the short times involved in these experiments, detailed, quantitative measurements of the self-luminous flow field behind the body have been difficult to obtain (2). Also, visualization of the flow field by ordinary shadowgraph and schlieren techniques has been difficult at the low densities required for simulation (3). The techniques described in this paper fall into two general categories: (a) photoelectric instruments for measuring the radiation intensity and its distribution in the wake, and (b) instruments for visualizing (photographing) the flow field. The instruments are described, typical data are shown, and some of the information available from the data is discussed. Other papers (4), (5), (6) illustrate the importance of these data in guiding the understanding of wake phenomena.

The experiments described herein were carried out either in the AERL ballistic ranges, using arc-heated light gas guns, or at the Canadian Armament Research and Development Establishment (CARDE) ballistic range facility. * The models were unsaboted plastic spheres of 0.22, 0.55 and 0.88-inch-diameter fired at Mach numbers of 15-25. In most cases air at controlled

*The authors wish to thank Mr. George Tidy, Superintendent, Aero/Physics Wing, and Dr. Andre Lemay, Section Leader, Radiation Physics Section, for their assistance and cooperation in the CARDE program.

density and ambient temperature was the test fluid. For some experiments, argon, nitrogen, or xenon were used to change the flow conditions or to increase the optical radiation. The optical radiation observed in these experiments is due predominately to ablative impurities in the viscous core of the wake (7).

PHOTOELECTRIC INSTRUMENTS

Photoelectric Recorder

Two photoelectric techniques have been devised for obtaining quantitative measurements of the radiation intensity from selected volume elements of the self-luminous hypersonic wake. In the simpler of these instruments, the photoelectric recorder, the radiation from the projectile and its wake is monitored by a photomultiplier tube looking through a thin slit oriented perpendicular to the flight direction. A schematic of this instrument is shown in Figure 1. A slit, positioned in front of the photomultiplier tube, is imaged into the range by a spherical mirror. The dimensions of the slit and the optical magnification control the area in the test section from which the radiant intensity is integrated by the instrument. At AERL, a 10" focal length mirror is used to give an optical magnification of about one. Thus the image size is nearly the slit size, typically 1.0" x 0.020". The photomultiplier used has S-11 response giving it a spectral sensitivity from 3200-6200 Å. No filters have been used in obtaining data discussed here. The electronics and oscilloscopes employed with this instrument are sufficiently fast that the response of the instrument is the transit time of events across the slit—approximately 0.1 μ sec for a projectile flying at 15,000 ft/sec.

A sample oscillogram from the photoelectric recorder for a firing of a 0.22-inch-diameter nylon sphere into air at a pressure of 7.0 cm Hg is shown in Figure 2. The projectile crosses the slit about 6 μ sec after the scope has been triggered. The first sharp rise in intensity results from the stagnation point radiation. The gas cools as it expands around the body, and the radiation intensity falls. A few microseconds later, a second but less abrupt rise in intensity occurs. The timing of this rise is consistent with the position of the recompression shock as seen in schlieren or shadowgraph photographs. Following this second peak, the radiation gradually decays downstream in the wake. The fluctuation in intensity associated with the decay can be correlated with turbulent eddies in the wake. The photoelectric recorder, together with the streak drum camera technique discussed below, achieved the first experimental evidence of transition from laminar to turbulent flow in the viscous core of the hypersonic wake (4).

Wake Scanner

The second of the photoelectric techniques developed at AERL is the wake scanner. This instrument makes it possible to repetitively scan the wake and obtain profiles of radiation intensity across the flow field, i.e., in a direction perpendicular to the flight path. The instrument is shown schematically in Figure 3. The optical system images the 17 fixed mirrors, each of which is masked to form a rectangular slit, into the range. As the turbine-driven

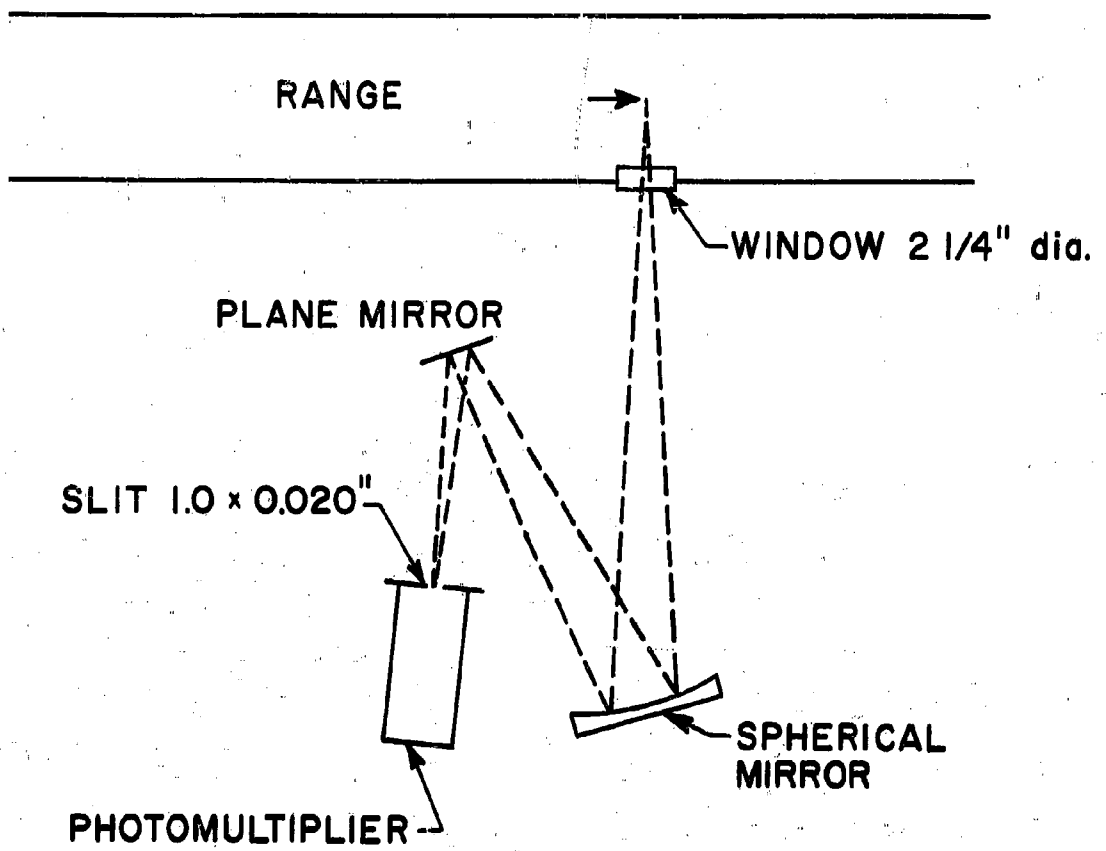


Fig. 1 Schematic of photoelectric recorder. View is from the top. The arrow in the range indicates a model moving from left to right across the test station. The dimensions shown here are for the AERL range.

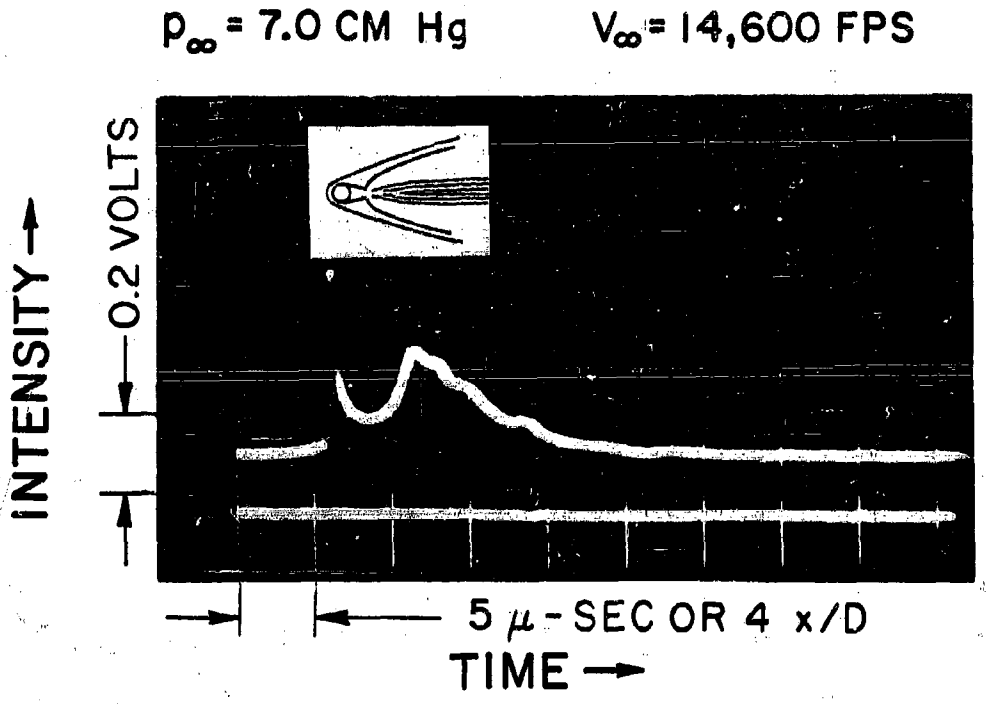


Fig. 2 Sample oscilloscope record of photoelectric recorder output. The sketch at the top of the photograph has been drawn approximately to scale to aid in correlating the recorder signal with the flow field events. The data were taken on the AERL range firing 0.22-inch-diameter nylon spheres. Note that the horizontal scale, time, may be converted into distance downstream of the stagnation point in body diameters, x/D.

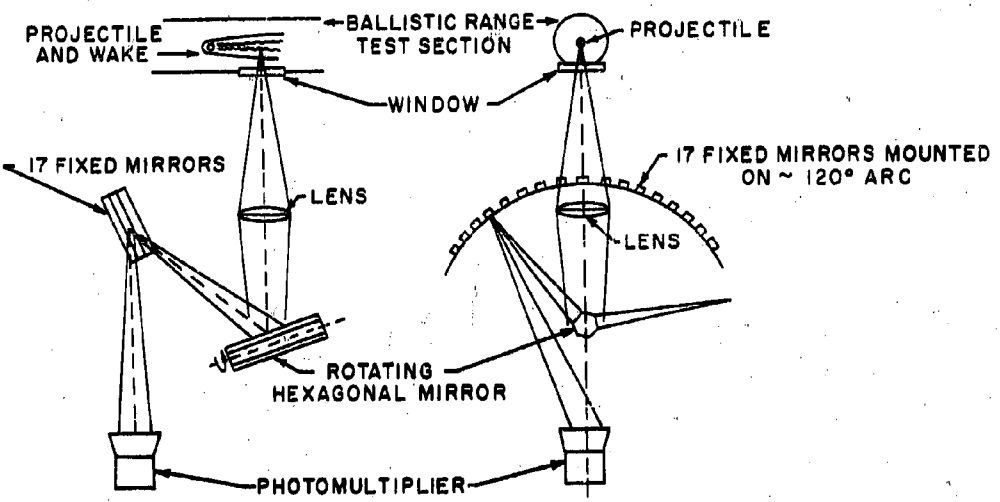


Fig. 3 Schematic of wake scanner instrument. The left-hand view is from the top, the right-hand view from the side.

hexagonal mirror is rotated the image of each slit is moved across the range from top to bottom. The intensity of radiation seen by an image as it traverses the range is monitored by a photomultiplier tube and displayed on an oscilloscope. A photomultiplier tube with S-11 response is also used in this instrument. The wake scanner is designed so that the 17 fixed mirrors are seen sequentially by each of the six faces of the rotating mirror. Thus 102 equidistant scans across the wake are made per revolution of the rotating mirror. At a mirror speed of 1500 revolution/sec* a scan of the wake can be made every 6.5 μ sec. Increasing the scanning frequency may be achieved by increasing the mirror speed. Conversely, the scanning frequency may be decreased by masking off intermediate fixed mirrors or decreasing the rotational speed of the mirrors.

As used on the AERL range (firing 0.22-inch projectiles) with a 7-inch focal length, f/2.5 lens, the image of each slit in the range is typically 1/4-inch long by 1/16-inch high. Therefore, normal to the wake axis in the scanning direction the instrument resolves about 1/4th of a body diameter, while its resolution axially is about one diameter. The distance between images is about 1.2 inches (measured normal to the wake axis). This distance can be increased at a sacrifice in the number of scans per mirror revolution by masking off some of the fixed mirrors. With the conditions listed above and with the rotating mirror driven at 1500 revolutions/sec, the linear velocity of each image traversing the range is about 15,000 ft/sec.

A sample of data from the wake scanner is shown in Figure 4 from a firing of a 0.22-inch-diameter nylon sphere into argon at a pressure of 2.0 cm Hg. In the left-hand oscillogram the upper beam displays the signal from the photoelectric recorder while the lower beam records the wake scanner output. The two instruments are observing the same position in the range. The display of the photoelectric recorder is seen to be qualitatively somewhat different than that in Figure 2. This is a result of the facts that the test gas in Figure 4 is argon, and that the wake is laminar under these flight conditions. It can be seen that as the model passes the test station the wake scanner records a large signal and drives the oscilloscope off scale. The sensitivity of the instrument has been set high in order to observe radiation in the far wake and, therefore, details around the body have been lost. The instrument soon recovers, and scans across the wake are obtained about every 7 μ sec. On the right-hand oscillogram three of the scans are presented at a faster sweep speed illustrating the detail available by this technique. An interesting feature of this data is that it can be interpreted to indicate that the laminar wake is either a luminous axisymmetric shell with a darker center core or a multiple filament structure.

Besides this interesting qualitative feature, quantitative results can be obtained from this data. For example, from pictures such as Figure 4 the

*The rotating mirror is manufactured by Avco Research and Advanced Development Division, Wilmington, Mass. and can be operated up to 3000 revolutions/sec.

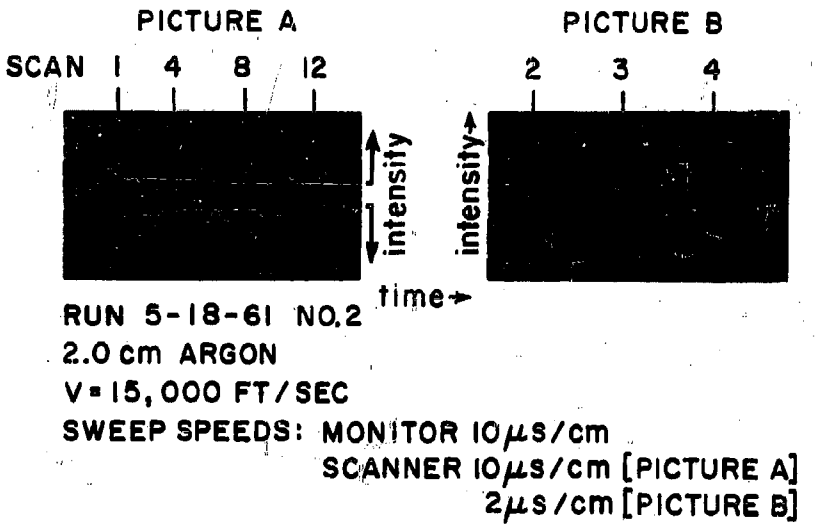


Fig. 4 Typical wake scanner data for the firing of a 0.22-inch-diameter nylon sphere into argon. Picture A shows both a photoelectric recorder signal (top beam) and the wake scanner output (lower beam). Note that the sense of increasing signal (radiation intensity) is opposite for these two signals. Picture B displays three wake scans in greater detail.

width of the luminous wake and its growth downstream of the projectile can be determined. Such information is now available for a number of runs in air, argon, and xenon for a range of free stream pressures and velocities. Wake width measurements made with the wake scanner have been compared with such data from other techniques; e.g., shadowgraph and race track pictures.

A reconstruction in projectile coordinates of the wake scanner data is shown in Figure 5. The intensity profiles obtained in the individual scans are displayed at their proper positions downstream of the model. From such a reconstruction the real time development of the luminous wake can be observed. The results in Figure 5 are for a firing of a 0.22-inch-diameter lexan sphere into xenon at 1.0 cm Hg pressure; for such conditions the wake is turbulent. The asymmetric pattern of the intensity profiles is indicative of large scale turbulent motion in the luminous wake.

FLOW VISUALIZATION TECHNIQUES

Streak Drum Camera

In early attempts at AERL to visualize the wake by its own luminosity, a streak drum camera (8) was used to photograph the flow field at low pressures. The experimental arrangement is shown in Figure 6. The drum which carries the film is pneumatically driven and rotates the film at a velocity of up to 0.1 mm/μ sec at the focal plane. The optical system of the camera (f/2.5) images the motion of the model and its wake on the film. As the model and its wake move down the range, their image moves transverse to the motion of the film, producing on the film an x-t diagram of the flow field events. The film used in these studies was Royal-X-Pan developed to obtain the highest film speed.

Figure 7 shows typical streak drum camera records for nylon spheres fired into argon. Figure 7 (a) is an x-t diagram to aid in interpreting the pictures. The bright diagonal streak across each photograph is caused by the luminosity in the stagnation point region. The luminous flow field shows up as a family of curved streaks in Figure 7 (b). These streaks are interpreted as being luminous, large scale turbulent eddies which start out with the model velocity and slow down as the gas expands and flows back into the wake. In Figure 7 (c) the wake appears as luminosity without structure. This is interpreted as a laminar wake. The two vertical dark lines at the left in each picture are calibration masks placed on the window about one inch apart. Each picture shows about 18 inches of flight path. The drum camera photographs, such as Figures 7 (b) and (c), together with the photoelectric recorder data described earlier are used to differentiate between laminar and turbulent wakes (4).

From photographs such as Figure 7 (b) the velocity of the luminous eddies can be determined. Such data are shown in Figure 8 where the eddy velocity in a coordinate system fixed in the model, V_s, normalized by the model velocity, V_∞, has been plotted against distance downstream in the wake, for several pressures of air and argon. It can be seen that for air

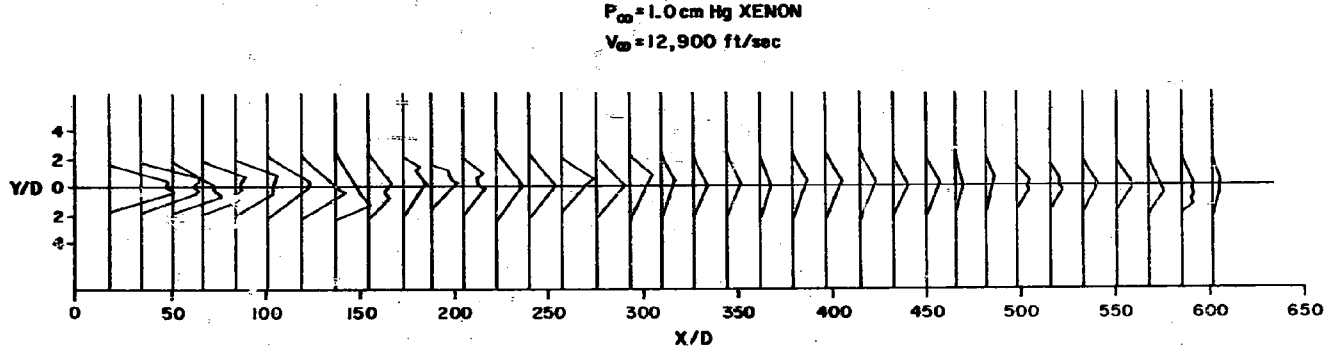


Fig. 5 Display of wake scanner profiles for a firing of a 0.22-inch-diameter lexan sphere into xenon. The individual intensity profiles have been placed along a scale of distance downstream in the wake from the stagnation point in body radii, X/R_N. The vertical scale gives the width of the profiles in body radii, Y/R_N.

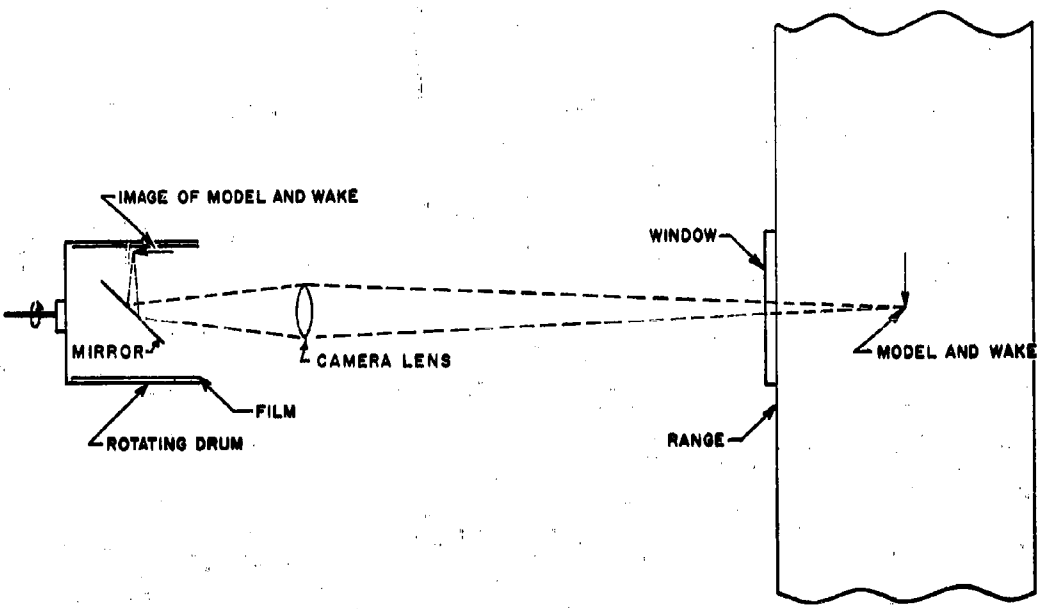


Fig. 6 Schematic of drum camera. View is from the top. The arrow (projectile and wake) is flying from top to bottom in the sketch while its image on the film moves from right to left.

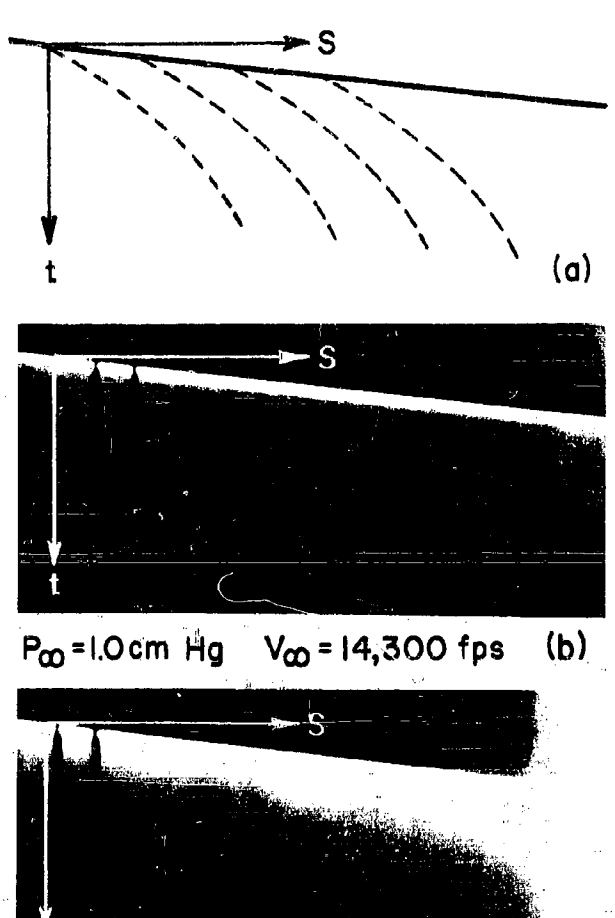


Fig. 7 Typical streak drum camera data. (a) An s-t (distance-time) diagram to aid in interpreting the photographs. The diagonal heavy line represents the projectile and stagnation point radiation. The dashed, curved lines represent luminous eddies which start out with the projectile velocity but, as they flow downstream in the wake, slow down eventually coming to rest. (b) and (c) Streak drum camera photographs of 0.22-inch-diameter nylon spheres fired into argon. In (b) the luminous streaks are interpreted as luminous large scale disturbances in a turbulent wake. In (c) the diffuse, structureless luminosity of the wake is interpreted as caused by a laminar flow field.

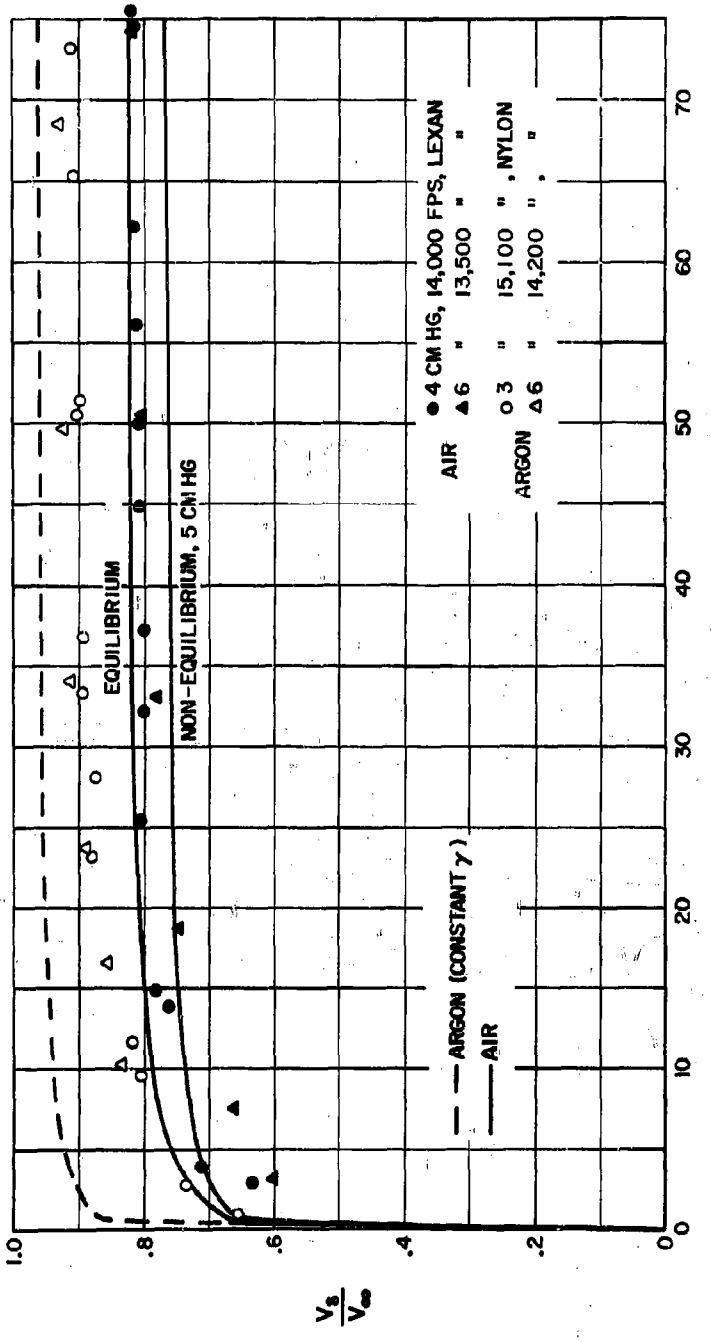


Fig. 8 Eddy velocity as obtained from drum camera photographs, such as Fig. 7 (b), for firings of 0.22-inch-diameter spheres into air and argon. The eddy velocity, V_s, is given in the steady state (body fixed) coordinates, normalized by the model velocity, V_∞. The solid curves are theoretical calculations which show good agreement with the experimental data.

the velocity of the turbulent eddies has decayed to about 0.2 of the model velocity as close as 15 body diameters downstream in the wake. The solid lines on Figure 8 are theoretical calculations of the gas velocity in the wake at the edge of the viscous core, as discussed in Reference (4). Besides yielding information on eddy velocity in the wake, these drum camera pictures can also be analyzed for the shedding frequency of the eddies. These data have been useful in drawing comparisons between the wake behind hypersonic models and the wake behind bluff bodies in incompressible flow (6).

Race Track Flow Visualization Technique

The race track method is a technique of flow visualization (8), (9) which enables self-luminous events in the wake to be photographed when the illumination is too low and the motion is too fast for conventional high speed movies and snapshots. The basic principle of the method is as follows: A drum camera is used to form an image of the object on a moving film strip. The orientation of the moving image and the rotating film is 90° different from the streak drum camera technique described previously, in that the film is moved parallel to the projectile motion. By adjusting the drum speed to match the image speed, the action is effectively stopped and longer exposures may be employed without blurring the image. The exposure is determined by the time, t_e, it takes the object to cross a slit placed in front of it, perpendicular to its motion. Of course, tracking is possible in only one direction and some resolution is lost if the event to be photographed does not move parallel to the film motion. The amount of blurring which results from miss-tracking is $(\underline{V}_i - \underline{V}_f)\, t_e$, where $\underline{V}_i$ is the vector velocity of the image and $\underline{V}_f$ is the vector velocity of the film. For the hypersonic wake the radial motion is expected to be a small fraction of the axial motion, and, therefore, the loss in resolution resulting from miss-tracking can be kept small. In any event the resolution and the exposure are controlled by the slit width.

A schematic diagram of a race track set-up is shown in Figure 9. It consists of (a) an f/2.5 drum camera, (b) a row of 3 vertical slits 0.2 inches wide mounted on the range window 0.67-inch apart, and (c) an array of mirrors to preserve a constant optical path from each slit to the film. The purpose of the multiple slits is to produce several images for studying the spatial correlation of luminous events as they move downstream in the wake. The array of mirrors displaces vertically the image from each slit so that the successive images do not write over each other on the film. A different arrangement of the slits and optical system is used to allow orthogonal viewing of the wake. By using slits of different widths, the dynamic range over which the luminosity can be recorded may be greatly increased.

In the present work the system was adjusted to track luminous events far downstream of the model. The wake velocity was determined from streak drum camera data such as Figure 8. For projectile velocities of about 15,000 ft/sec, the effective exposure times were about 5 μ sec. Since a tracking accuracy of about 10% was achieved, the image quality is the same as that which would be obtained in a snapshot at one-tenth the exposure.

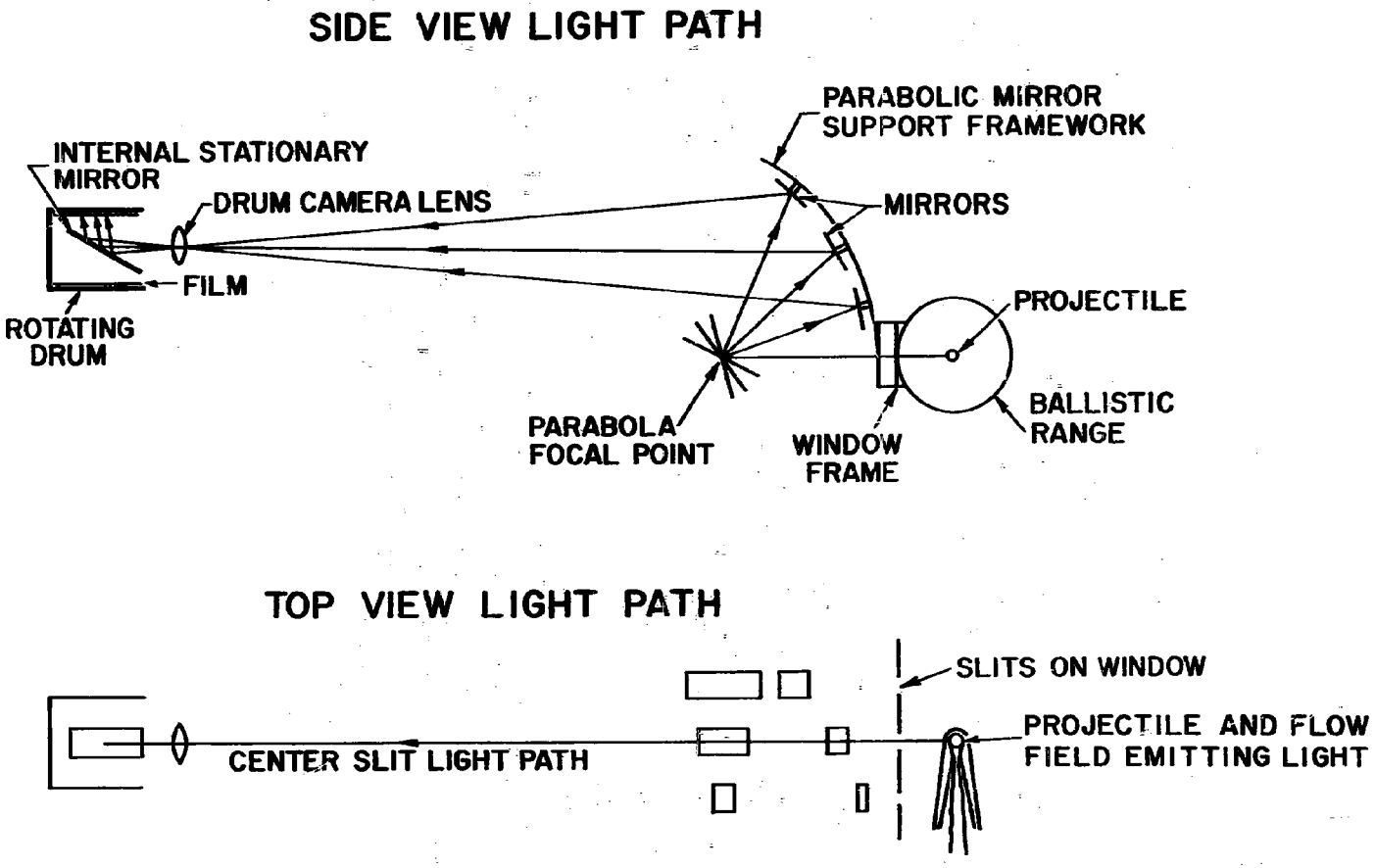


Fig. 9 Schematic of the race track flow visualization technique. In this arrangement the motion of the luminous wake as seen through the slits is tracked by the film on a rotating drum.

Some typical race track photographs are shown in Figures 10 and 11. Figure 10 shows photographs taken at AERL of 0.22-inch-diameter lexan spheres fired into xenon and argon. Both turbulent and laminar wakes are shown. The pictures in Figure 11 were taken at CARDE firing 0.55-inch-diameter lexan spheres into air. The development of the turbulent wake as the free stream pressure (Reynolds number) increases is evident. Note that the region close to the pellet where the difference between the gas velocity and film speed is large is not tracked and no detail is evident in this portion of the picture.

In interpreting these photographs it is important to keep in mind that they are not snapshots. As in the case of focal plane shutters, different parts of the picture are exposed at different times. This leads to a distortion of the photograph in which, although the apparent length of the wake is shortened, the structure of the events which are being tracked is correctly reproduced. The scales shown in Figure 10 give the distance between the projectile and an event at the instant that the event is photographed. However, that scale does not represent the true distance between events in the wake (9).

Besides providing information about the development of the turbulent wake these race track photographs can be analyzed for the growth of the luminous wake. In this respect the technique complements the wake scanner. Orthogonal race track photographs have also yielded evidence of asymmetries in the hypersonic wake (6).

Sodium Schlieren

In attempting to use schlieren photography for the study of transition from laminar to turbulent flow in the wakes of high speed projectiles, it has been found necessary to reduce the gas density to levels close to the limits of sensitivity of current techniques. Under these conditions, when one looks at a set of photographs of decreasing pressure, it is exceedingly difficult to determine whether the disappearance of the characteristic pattern of small scale turbulence in wakes results from laminar flow or from loss of sensitivity. The present work was undertaken in an attempt to remove this ambiguity by achieving at least an order of magnitude improvement in schlieren sensitivity (10).

At the time the work was begun, it was felt that refinement of the optics had already been pushed to the point of diminishing returns. Since the sensitivity of a schlieren system is directly proportional to the refractivity of the test medium, it was decided to explore the possibility of enhancing the refractivity of the gas by using radiation close to a resonant transition. One advantage of enhancing the refractivity is that spurious refractive variations in the windows become less important. Calculations using the dispersion formula of classical radiation theory (11) show that the absolute value of refractivity of a gas close to a resonance line can be many orders of magnitude larger than that of air in the visible portion of the spectrum, far from its resonance in the vacuum ultraviolet. Since the technical difficulties associated with working in the vacuum ultraviolet are very substantial, it appeared more attractive to attempt enhancement of the refractivity by using a gas with strong resonances in the ultraviolet or visible region. Sodium vapor with

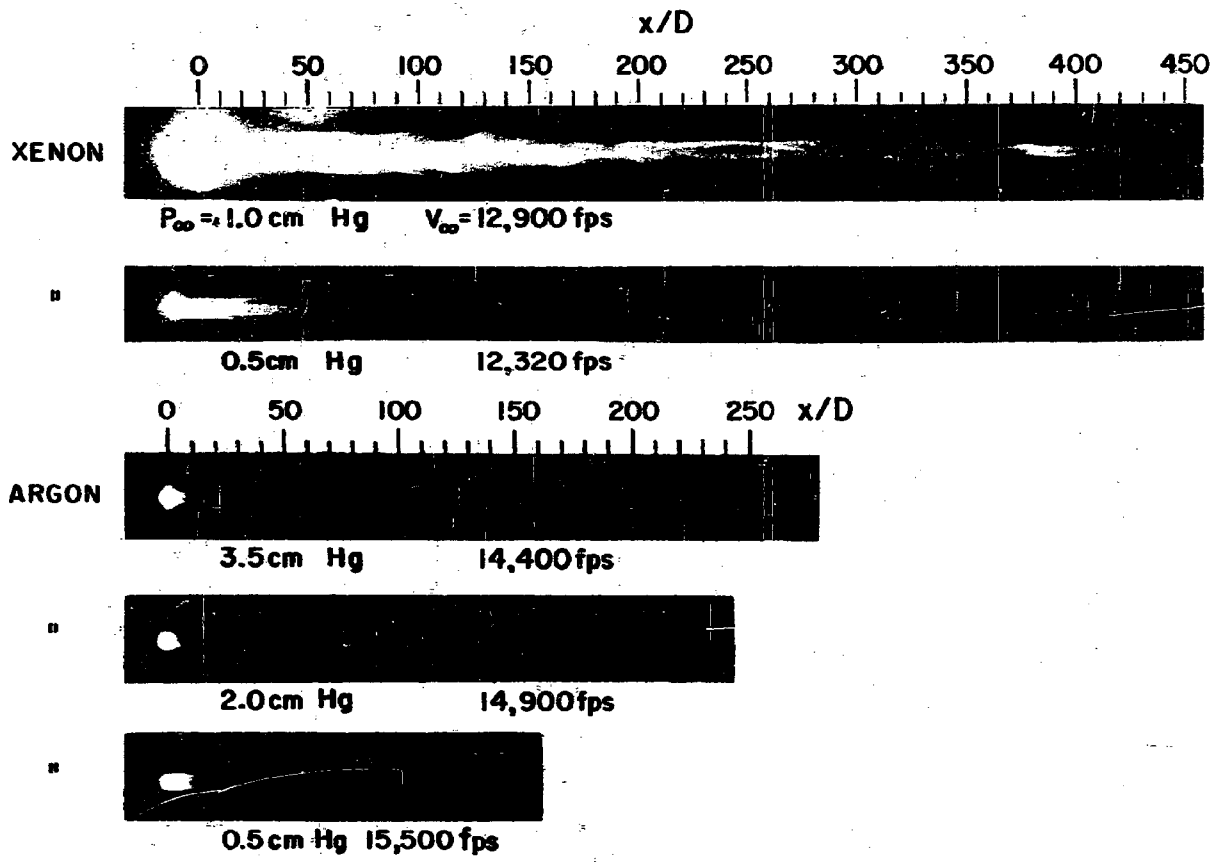


Fig. 10 Typical race track photograph. The top two pictures are for 0.22-inch-diameter lexan spheres fired into xenon. The large-scale turbulence of the top photograph is to be compared with the laminar flow of the picture beneath it at a factor of two lower pressure. The lower three photographs are for nylon spheres fired into argon. Although the radiation intensity is lower than for the xenon runs, the character of the flow field is evident from the pictures.

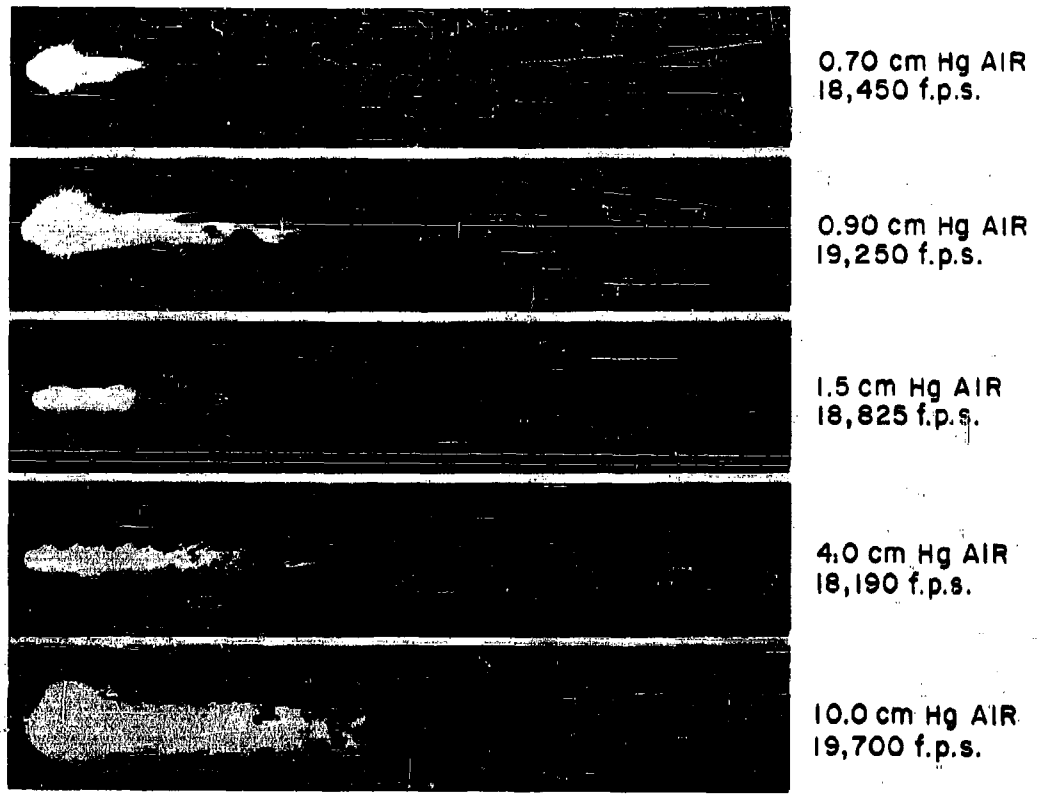


Fig. 11 Race track photographs of 0.55-inch-diameter lexan spheres fired into air for a range of free stream pressure. The development of the turbulent wake as the pressure (Reynolds number) is increased is evident.

its strong resonance lines near 5896 Å seemed the most convenient choice for an initial experimental study. The radiation is in the visible range of the spectrum, so that standard light sources and schlieren instrumentation can be used, and for which appropriate narrow band filters are readily available.

The absolute value of the refractivity of sodium (12) as a function of pressure for various wavelengths close to its resonance lines is compared with that of air in Figure 12. It can be seen for example, that if one operates with light 20Å from resonance one obtains with sodium vapor a refractivity more than two orders of magnitude larger than is obtained with air at the same pressure.

A schematic of the apparatus used in the present work is shown in Figure 13. The optical system is the usual schlieren off-axis parabolical mirror arrangement with a narrow band interference filter placed just in front of the knife edge. The remainder of the apparatus consists of an oven constructed of two 30" lengths of 2" diameter pipe in the form of a cross which is placed inside the test section of the ballistic range. The central 6" of the cross can be heated to temperatures of 500°C to maintain the required partial pressure of sodium vapor (Figure 12). The arms of the cross are water-cooled to keep the sodium vapor from condensing on the windows. Sodium is placed in the center of the cross and the total pressure is fixed with the working fluid of either argon or nitrogen.* The absolute vapor density of sodium is determined as a function of wall temperature by measuring the percent transmission through the vapor of light selected by an interference filter and comparing with similar test cell transmission calibrations (12). The sodium level reached in the central region has been observed to be close to the equilibrium vapor pressure for the measured wall temperature.

Figure 14 shows six wake photographs taken using a filter 70Å wide with the peak transmission 20Å toward the red from the resonance line, and an exposure of one microsecond. These are photographs of 0.22-inch-diameter nylon spheres fired into nitrogen with 0.1% of atmospheric density of sodium vapor added. The three pictures in the right-hand column, taken at about the same velocity, represent a transition sequence as the free stream density (Reynolds number) is increased. The photographs at constant pressure (top row, middle row and lower left) indicate decreased structure as the free stream velocity increases. Sodium schlieren pictures have now been obtained at densities of 0.51 to 10.0% atmospheric and at velocities between 1,300 and 13,000 ft/sec for 0.22-inch-diameter nylon spheres. The results of this study are contributing to the understanding of the aerodynamics of the wake (5).

*Air cannot be used with sodium vapor at these temperatures because of the reaction of sodium with oxygen.

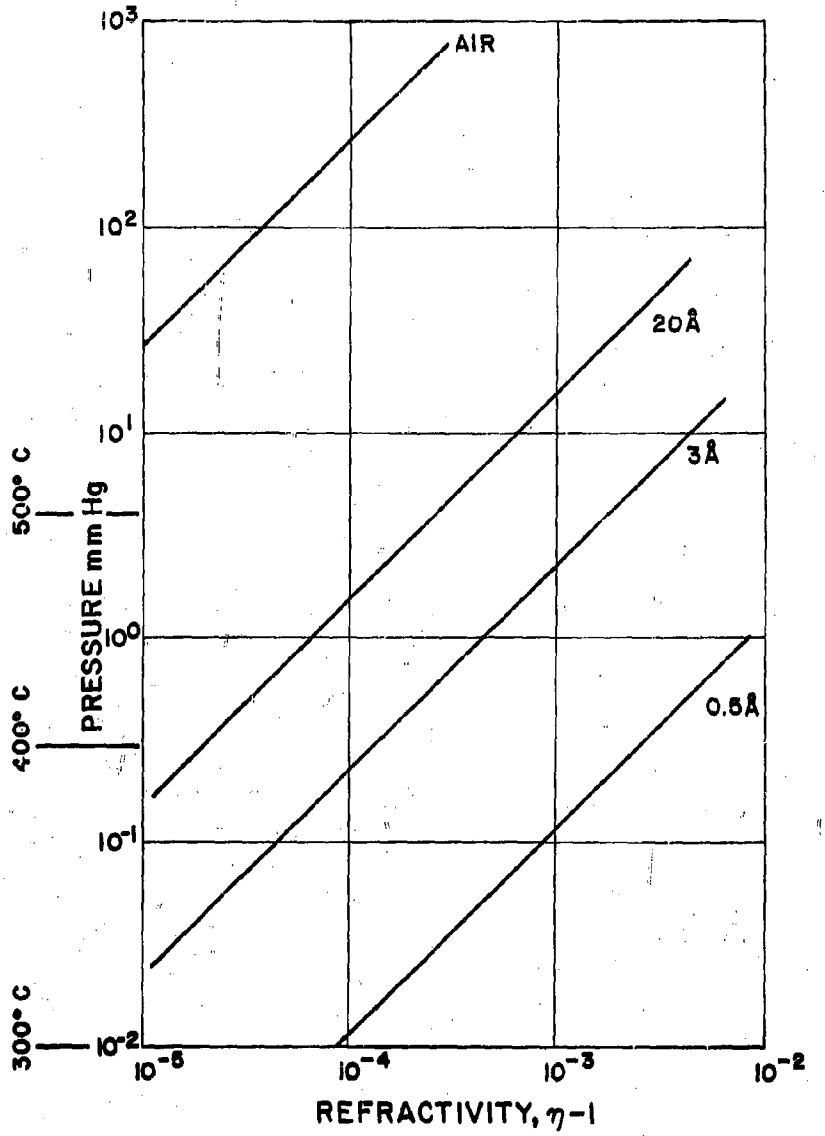


Fig. 12 Refractivity of sodium vapor and air as a function of pressure. The three curves labeled 0.5Å, 3Å and 20Å are the refractivity of sodium vapor measured with light 0.5Å, 3Å and 20Å, respectively, from the resonance line at 5896Å. The temperature scale indicates the heating necessary to produce the corresponding equilibrium vapor pressures of sodium.

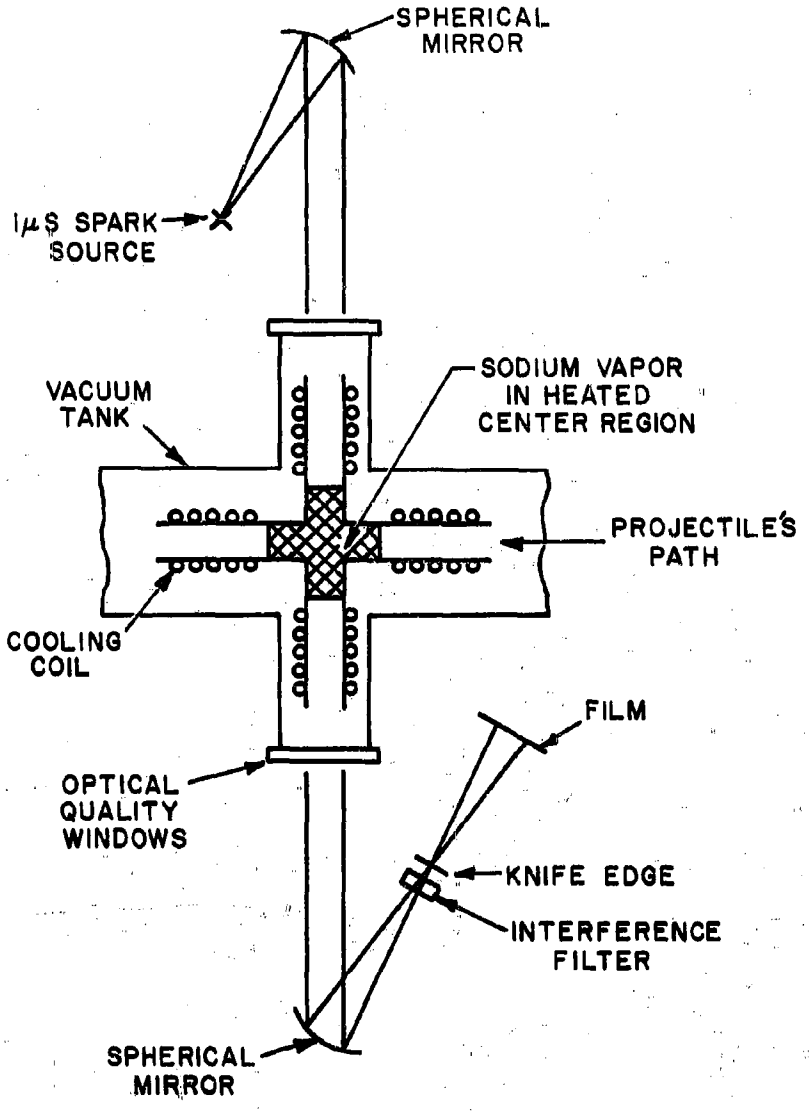


Fig. 13 Schematic of sodium schlieren apparatus. The cross in the center of the system is an oven to produce a controlled partial pressure of sodium vapor in the working fluid.

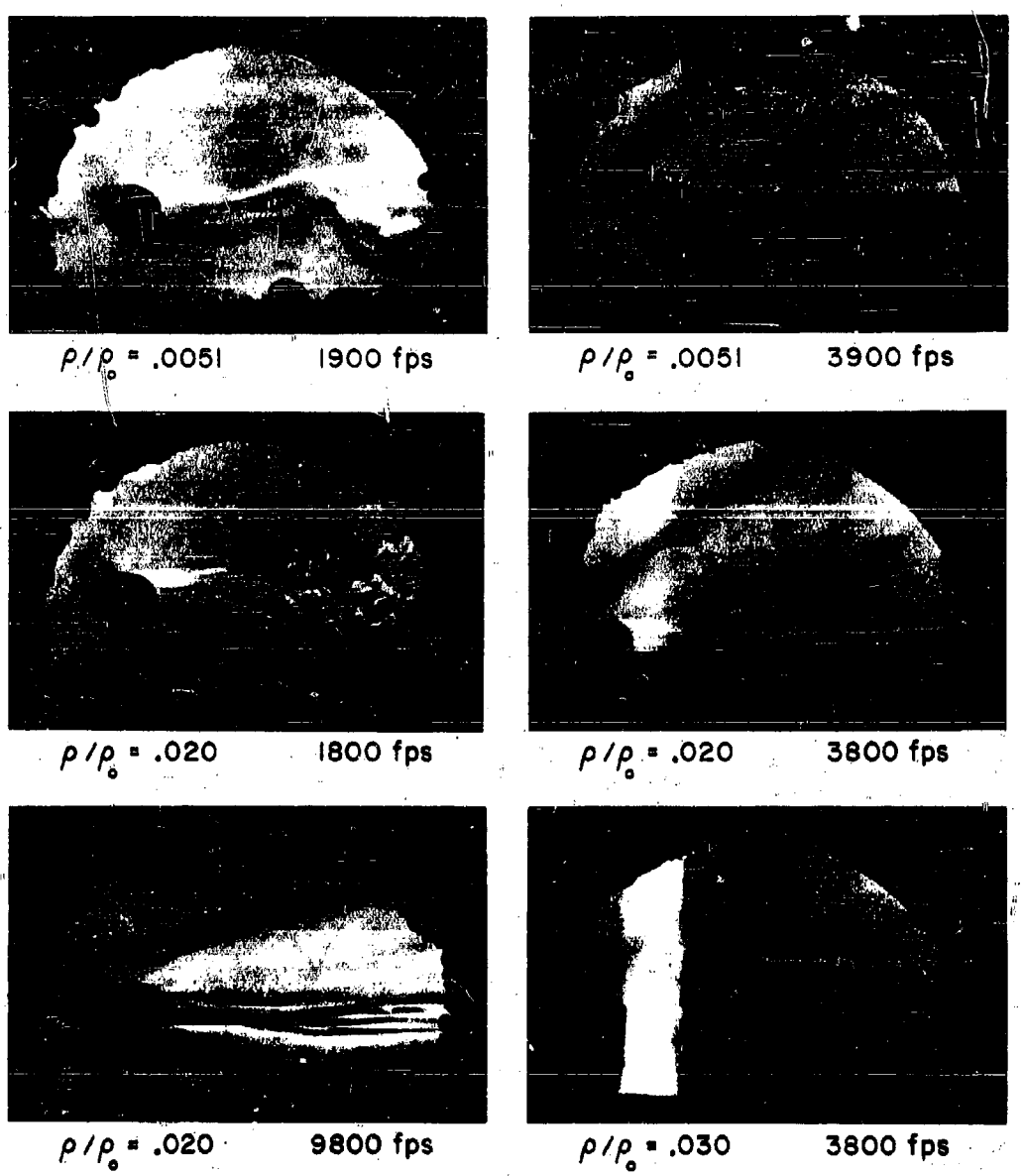


Fig. 14 Schlieren photographs of 0.22-inch-diameter nylon spheres fired into nitrogen at indicated density with 0.1% ρ/ρ_o of sodium vapor added. The free stream speed of sound under these conditions is about 2000 ft/sec, so that the top two pictures at the left are subsonic.

Image Converter

Although the low levels of luminosity and short exposure times necessary for taking snapshots of the hypersonic wake have led to the development of the race track technique discussed previously, the recent development of the image converter camera has brought about re-examination of the direct photographic technique. The image converter camera (13) is an instrument in which both electronic light amplification and fast shuttering are available. An image converter camera* was recently purchased by the laboratory and installed on the AERL range to photograph the base region of the flow field. This instrument utilizes an RCA tube with S-11 response and has a light gain of about 50 from the photocathode to the photoanode. A framing unit is available which allows three consecutive exposures to be made with independently adjustable intervals between frames. The frames are electronically deflected on the photoanode face to prevent double exposures.

Photographs by this technique for 0.22-inch-diameter lexan spheres are shown in Figure 15 for three gases air, argon, and xenon. In all cases the free stream pressure is sufficiently high so that the wakes are turbulent. Considerable detail is evident in these pictures, and it is hoped that this data may provide information on the mechanism of vortex shedding and the onset of turbulence in the wake.

SUMMARY

This paper has discussed a variety of techniques that have been developed for obtaining data on the wake behind hypersonic projectiles in the ballistic range. Photoelectric instruments, the recorder and wake scanner, are now available for measuring the intensity of radiation from selected volume elements of the wake. The streak drum camera and its race track adaptation allow the self-luminous flow field to be visualized. The sensitivity of the schlieren photographic technique has been increased by the use of resonance radiation to the point where pictures showing wake structure at densities less than 1% of atmospheric density can be obtained. The development of the image converter camera has made it possible to take snapshot photographs of the self-luminous flow field with exposure times sufficiently short that details of the base flow region are now available.

The types of wake data available from these instruments are: (a) rate of decay of radiation, (b) transverse profiles of radiation, (c) eddy velocity history, (d) eddy shedding frequency, (e) width and growth of the luminous wake, (f) Reynolds number for transition from laminar to turbulent flow, and (g) statistical structure of the turbulent wake. These data are having a significant effect upon the development of the understanding of the wake behind blunt bodies flying at hypersonic Mach numbers.

*Model C Image Converter Camera manufactured by STL Products, Inc., Los Angeles, Calif.

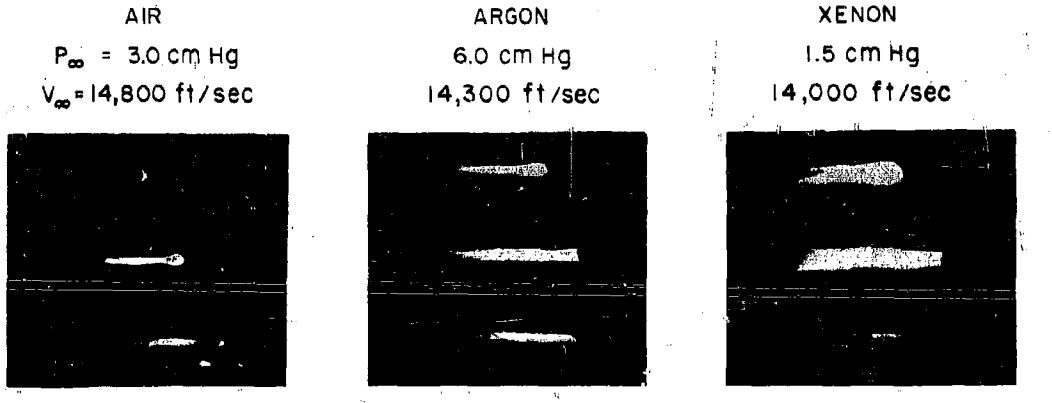


Fig. 15 Image converter photographs of the base flow region of 0.22-inch-diameter lexan spheres in air, argon and xenon. There are three frames per picture, the time sequence running from top to bottom. The projectile is moving from left to right in each photograph. The range conditions are given at the top of the pictures and the image converter camera conditions at the bottom.

REFERENCES

1. Bull, G. V., "Re-entry Studies in Free Flight Ranges," IAS Reprint No. 59-143 (1959).

2. St. Pierre, C. A., "Results from the Third Radiation Program in CARDE Hypersonic Range 1, Using 0.40 Inch Diameter Spheres," CARDE Technical Memorandum 611/61, Canadian Armament Research and Development Establishment, June 1961.

3. Slattery, R. E. and Clay, W. G., "Width of the Turbulent Trail Behind a Hypervelocity Sphere," The Phys. of Fluids, vol. 4, p. 1199 (1961).

4. Hidalgo, H., Taylor, R. L. and Keck, J. C., "Transition from Laminar to Turbulent Flow in the Viscous Wake of Blunt Bodies Flying at Hypersonic Mach Numbers," Avco-Everett Research Laboratory Research Report 133, April 1961; (accepted for publication in J. Aero/Space Sci.).

5. Goldburg, A. and Fay, J. A., "Vortex Loops in the Trails Behind Hypervelocity Pellets," Avco-Everett Research Laboratory AMP 75, February 1962.

6. Fay, J. A. and Goldburg, A., "Vortex Shedding, Transition and Growth of Luminous Wakes Behind Spheres at Mach 15 to 25," Avco-Everett Research Laboratory Research Note 295, May 1962.

7. Lemay, A., "Radiation Measurements from the Plasma Sheath Surrounding Hypersonic Projectiles," CARDE Technical Memorandum 693/62, Canadian Armament Research and Development Establishment, March 1962.

8. Rinehart, J. S., Allen, W. A. and White, W. C., "Phenomena Associated with the Flight of Ultra-Speed Pellets, Part III, General Features of Luminosity," J. Appl. Phys., vol. 23, p. 297 (1952).

9. Washburn, W. K. and Keck, J. C., "The Race Track Flow Visualization of Hypersonic Wakes," Avco-Everett Research Laboratory Research Report 131, March 1962; (accepted for publication in ARS Journal).

10. Leonard, D. A. and Keck, J. C., "Schlieren Photography of Projectile Wakes using Resonance Radiation," Avco-Everett Research Laboratory Research Report 130, March 1962; (accepted for publication in ARS Journal).

11. Heitler, W., "The Quantum Theory of Radiation," 3rd ed. Oxford: Clarendon Press, (1954).

12. Wood, R. W., "Physical Optics," 3rd ed. New York: Macmillan, (1934).

13. Chippendale, R. A., "The Photographic Efficiency of Image Converters," Proceedings of the Third International Congress on High-Speed Photography. New York: Academic Press, pp. 116-125, (1956).

BALLISTIC MISSILE DECOY OPTIMIZATION THROUGH DYNAMIC PROGRAMMING

Y. Fukuda
W. E. Faragher
L. L. Philipson
M. S. Schaeffer

Planning Research Corporation
Los Angeles 24, California

ABSTRACT

This paper presents a method for determination of optimum ballistic missile reentry decoy specifications through a dynamic programming model of the terminal battle. In this model the defense, given any fixed reentry decoy package specifications and discrimination decision rules, makes an optimized assignment of defense interceptor missiles (such as Nike Zeus) to the "apparent" warheads discriminated in a sequence of attacking missile clouds, so that the expected number of surviving warheads over the total number of clouds, E, is minimized. Then the defense determines the particular discrimination decision rules which minimize E, given the fixed offense decoy package. Finally, the offense selects the particular decoy specification which is optimal in that it maximizes, under payload size and weight constraints, the defense-minimized values of E. The optimal defense decision and interceptor assignment rules derived are real-time implementable. The optimized decoy package specifications are in terms of number, dimensions, weight, and discriminable parameters meaningful for design implementation.

BALLISTIC MISSILE DECOY OPTIMIZATION THROUGH DYNAMIC PROGRAMMING

Y. Fukuda, W. E. Faragher, L. L. Philipson, and M. S. Schaeffer
Planning Research Corporation
Los Angeles 24, California

I. INTRODUCTION

Consideration is being given to the advantages of containing in a ballistic missile payload reentry decoys, together with the warheads, in order to enhance penetration or survivability of the warheads against terminal defenses. Since there are weight and volume constraints for a payload of an offense missile, the number and size of decoys to be contained in a payload must be subjected to constraints and optimized design tradeoff possibilities therefore arise.

If the offense uses a relatively large number of decoys, then, assuming that all decoys are identical with respect to dimensions and weight, each decoy must be relatively small in dimensions and light in weight and therefore dissimilar to warheads. In this case, the defense may be easily able to discriminate among warheads and such decoys.

On the other hand, if a decoy is very similar to a warhead, its dimensions and weight then may require that there not be available a sufficient number of decoys to adequately dilute the defense interception capability.

Therefore the problem of this paper is the tradeoff establishing the best size and weight of a decoy to be used by the offense, as determined by an optimal specification of its discriminable parameters, and as constrained by size and weight payload limitations.

In this paper we take the viewpoint of a game of strategy between the offense and the defense, when the offense allocates a number, say P, of ballistic missiles to a single target which the defense is going to protect by employing a number, say M, of defense interceptor missiles in a Nike-Zeus-like system. A strategy of the offense is a selection of a decoy specification which, under size and weight constraints, uniquely determines the number, say N, of decoys to be contained in an offense missile payload together with a fixed number, say K (possibly greater than one), of warheads. To the defense a strategy signifies a set of discrimination decision rules for deciding among warheads and decoys and a mode of assignment of defense interceptor missiles.

We describe as the dynamic single target model the specific model of the defense discrimination and interceptor missile assignment against the $P(K + N)$ attacking objects in the terminal battle. Section II contains a

detailed exposition of the dynamic single target model. In particular, we make an application of dynamic programming to determine an optimal mode of assignment of the defense interceptor missiles for any given decoy kit specification.

When there is made a choice of an offense strategy, i.e., a decoy specification, and a defense strategy, i.e., modes of discrimination and assignment, there results from the dynamic single target model an optimal (i.e., max-min) expected number of surviving offense warheads. Section III describes the offense's optimized selection of a decoy specification. In principle we first let the defense choose a strategy that minimizes the expected number of surviving warheads against any given offense strategy, and then let the offense select the strategy (i.e., decoy specification) that maximizes under payload constraints these minimized expected numbers of surviving warheads.

II. DYNAMIC SINGLE TARGET MODEL

General Description

The dynamic single target model is concerned with a war game in which the offense allocates a number, say P, of ballistic missiles to a target that a Nike-Zeus-type system will defend by employing a number, say M, of defense interceptor missiles.* It is assumed to be known to the defense as well as to the offense that a payload of an offense missile contains a number, N, of decoys and a number, K, of warheads. The number N is uniquely determined by a choice of the offense concerning dimensions and weight of a decoy.

It is also assumed that a defense interceptor missile has a known kill probability, p_K, against an offense warhead it is assigned to. The offense's P missiles will successively arrive at the vicinity of the target, forming P successive clouds as targets of the defense interception. It is assumed that there are always exactly $L = K + N$ objects in each cloud which will be subjected to defense discrimination and assignment decisions.

If the defense discriminates an object as a decoy, then there will be no assignment of defense interceptor missiles to it. If the defense makes a decision that an object is a warhead as a result of discrimination, then the object will be assigned an appropriate number of interceptor missiles. (The problem of how many missiles to assign will be taken up later.) The defense's discrimination decisions are assumed to be independent, and no reassignment of interceptor missiles is considered. It is assumed that the defense will be able to make these discrimination and assignment decisions in a sequential manner, one at a time, in the order of arrivals of these L objects at a certain "final decision" altitude.

Thus the defense's discrimination problem in the dynamic single target model with P clouds involving PL objects may be considered as consisting

*No defense limitations due to other system components are assumed.

of PL decision problems in sequence, and the assignment problem of defense interceptor missiles in this model can be identified as a PL-stage decision problem. We will later exploit this formulation in applying dynamic programming to optimizing for the defense interceptor missile assignment rules.*

Discrimination Methodology

Let us now describe more quantitatively the discrimination procedure in the dynamic single target model. Generally, any number of discriminable characteristics, or parameters, may be used by the defense to discriminate among warheads and decoys. For illustrative purposes, however, a procedure to be described here is concerned with a case where only one discriminable parameter is used. We will later present a procedure that will be applicable to more general cases.

A basic assumption for the defense discrimination procedure in this model is that a measurement, or an observed value, of a discriminable parameter of an object taken by the defense is a random variable with a (one-dimensional) probability distribution.† This probability distribution in general may be uniquely determined by the dimensions and weight of an object, given the measurement system. Since there are two different types of objects, i.e., warheads and decoys, involved in a discrimination test, the defense is confronted with two different probability distributions.

It is also assumed for simplicity here that a discrimination test is based on a single "threshold." Suppose a certain discriminable parameter is chosen. Then the defense will set forth a threshold value, τ, on which a discrimination test with this parameter is to be based. If a measurement of the parameter of an object is greater than τ, the defense will classify the object as a warhead; if not, the object will be classified as a decoy.

Now that we have a decision threshold and the warhead and decoy probability distributions, we are able to calculate the probability, say p_{WW}, that a warhead will be classified as a warhead, and the probability, say p_{DW}, that a decoy will be classified as a warhead. These probabilities are illustrated by Figure 1. Intuitively, the probability p_{WW} is significant because a portion $1 - p_{WW}$ of warheads will pass the defense discrimination test as decoys on the average, and, since there will be no interceptor missiles assigned to them, they will automatically impact on the target.‡ The

*It will be realized later that these rules are established on an a priori basis and can be prestored in the defense system. The procedures considered are thus entirely feasible for real-time implementation.

†The randomness arises, e.g., from inherent fluctuations due to reentry phenomena and from measurement errors.

‡Of course, due to the nonunity kill probability of the defense interceptors, some of the warheads fired at will impact successfully also.

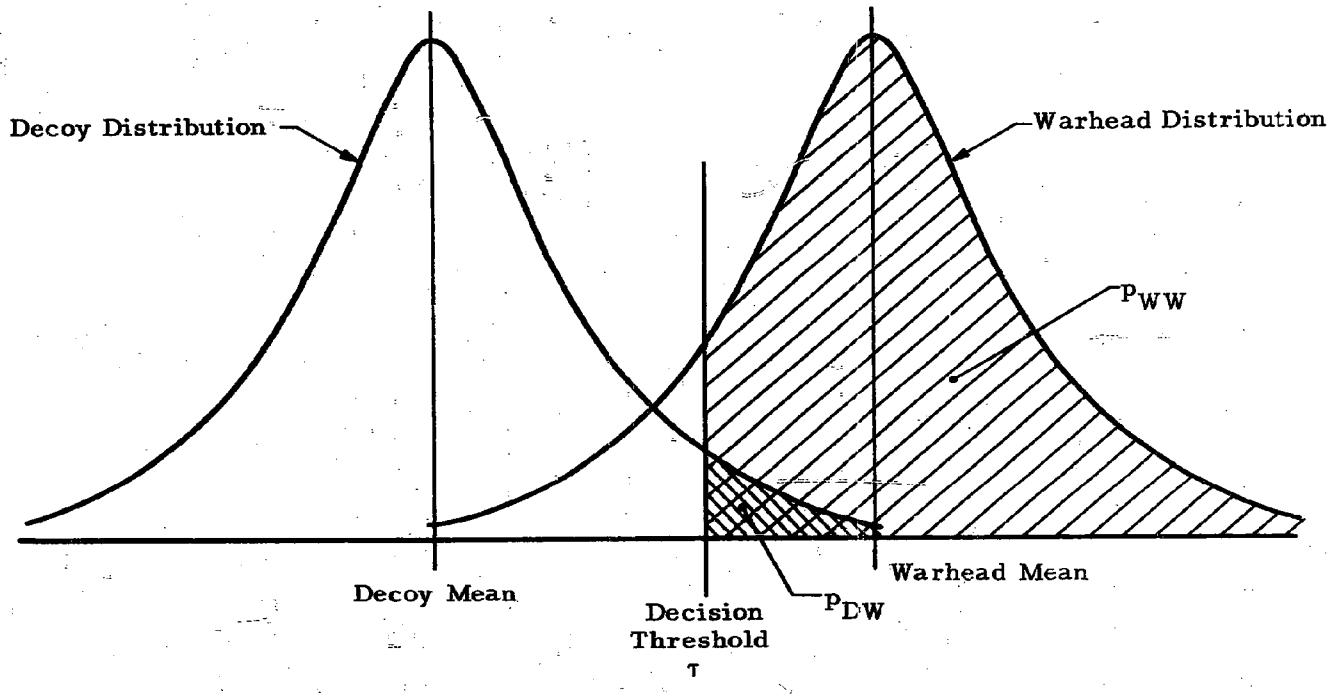


Fig. 1. Illustration of Defense Discrimination Test

significance of p_{DW} is that a p_{DW} portion of decoys will be assigned some interceptor missiles, and this results in their unavailability against later warheads.

We shall also employ the a priori probability, p_W, that a given object will be classified as a warhead by the defense discrimination procedure,

$$p_W = \frac{K}{L} p_{WW} + \frac{N}{L} p_{DW} \quad (1)$$

Assignment of Defense Interceptor Missiles

After a discrimination test, the defense will assign a number of interceptor missiles to an object that is classified as a warhead. A question still remaining to be answered is how many interceptor missiles the defense should assign to a given object. Since this assignment decision is made from object to object, or from stage to stage, as we mentioned before, we propose to apply dynamic programming for its optimized establishment.

The criterion of defense optimization in our formulation is to minimize the expected number of surviving warheads over the total PL decision stages, given any fixed decoy specification and discrimination thresholds. We identify the nth arriving object as stage n, so that the first object of the first cloud will be identified as stage 1, and the last object of the last cloud as stage PL. At each stage the interceptor assignment decision must generally take into account three factors: (1) the stage, or number of objects so far considered; (2) the number of these objects so far classified as apparent warheads in the present cloud; and (3) the number of remaining interceptors. In the following discussion we shall treat for explanatory purposes only the simplified case where information of type (2) is neglected by the defense. This means, admittedly somewhat unrealistically, that each assignment decision remains unbiased by whether many or few objects in the cloud have already been called warheads. The rather complicated modifications of the decision process and their implications to the decoy optimization procedures in the more realistic case are discussed in Appendix C.*

Let us now designate by $f_m(x)$ the minimum expected total number of surviving warheads over stages m through PL, when there are x interceptor missiles remaining available for these stages. The ultimate quantity of interest to us is $f_1(M)$, and to derive it we will have to determine recursively (and "working backwards") each $f_m(x)$ for $0 \leq x \leq M$, starting with $m = PL$.

Suppose we are at stage PL. Since this is the last object to be encountered by the defense, if it is classified as a warhead the defense can assign

*Since the modified procedure is a more powerful one for the defense, making use of more information (still in a real-time implementable manner), the results presented for the simplified case must be considered as somewhat optimistic for the offense.

all the remaining interceptor missiles to it. There may, however, be a restriction that at most t interceptor missiles can be assigned to any one object, t being a given number, say 3. If the last object is classified as a decoy, the defense will assign no interceptor missiles to it. To calculate the expected number of surviving warheads in this stage, i.e., $f_{PL}(x)$, we thus consider the following four mutually exclusive cases:

Case 1. Decoy is classified as warhead. The a priori probability that an object is a decoy is given by N/L, and a decoy will be classified as a warhead with a priori probability p_{DW}. But we are concerned only with the number of surviving warheads, and there will be no contribution to the expected number of surviving warheads from this case.

Case 2. Decoy is classified as decoy. A decoy will be classified as a decoy with probability $1 - p_{DW}$; but, as in Case 1, there will be no contribution from this case.

Case 3. Warhead is classified as warhead. The probability that an object is a warhead is given by K/L, and a warhead will be classified as a warhead with probability p_{WW}. The defense will assign x interceptor missiles to the object if $x \leq t$, and if $x > t$, the number will be t. The probability that a warhead will survive through an assignment of x interceptor missiles is given by $(1 - p_K)^x$, p_K being the kill probability of an interceptor missile. Therefore, the contribution to the expected number of survivors from this case is given by

$$\frac{K}{L}\, p_{WW} (1 - p_K)^z \tag{2}$$

where

$$z = \min (t, x) \ . \tag{3}$$

Case 4. Warhead is classified as decoy. A warhead will be classified as a decoy with probability $1 - p_{WW}$, and in this case no assignment of interceptor missiles will be made. Therefore, the warhead automatically surviving, the contribution from this case, is

$$\frac{K}{L} (1 - p_{WW}) \ . \tag{4}$$

Now $f_{PL}(x)$ is given by a sum of these contributions, that is,

$$f_{PL}(x) = \frac{K}{L} (1 - p_{WW}) + \frac{K}{L}\, p_{WW} (1 - p_K)^z \ . \tag{5}$$

We may rewrite Eq. (5) as

$$f_{PL}(x) = (1 - p_W) \frac{K}{L} \frac{1 - p_{WW}}{1 - p_W} + p_W \frac{K}{L} \frac{p_{WW}}{p_W} (1 - p_K)^z$$

and, defining $\tilde{p}_{WW}$ and $\tilde{p}_{DW}$ by

$$\tilde{p}_{WW} = \frac{K}{L} \frac{p_{WW}}{p_W} \tag{6}$$

and

$$\tilde{p}_{DW} = \frac{K}{L} \frac{1 - p_{WW}}{1 - p_W} , \tag{7}$$

we finally have

$$f_{PL}(x) = (1 - p_W) \tilde{p}_{DW} + p_W \tilde{p}_{WW} (1 - p_K)^z . \tag{8}$$

In terms of physical significance, $\tilde{p}_{WW}$ represents a probability that an object classified as a warhead is really a warhead, and $\tilde{p}_{DW}$ represents a probability that an object classified as a decoy is really a warhead. (These are "inverses" of p_{WW} and p_{DW}.)

Next, suppose we are at a stage m, $1 \le m < PL$, with x interceptor missiles remaining for the defense. If this mth object is classified as a decoy (probability of this event is $1 - p_W$), the defense will assign no interceptor missiles. Hence, applying probability $\tilde{p}_{DW}$, we may expect a surviving warhead from this object with probability $(1 - p_W)\tilde{p}_{DW}$, and there will still remain x interceptor missiles for stage $m + 1$, since none is used at stage m. If the object is classified as a warhead (probability of this event is p_W), and if the defense assigns y interceptor missiles to it, then, applying probability $\tilde{p}_{WW}$, we may expect a surviving warhead out of this object with probability $p_W \tilde{p}_{WW} (1 - p_K)^y$, and there will remain $x - y$ interceptor missiles for stage $m + 1$. Since our objective is to choose such a y that the expected number of surviving warheads from stage m onward may be minimized, we obtain an equality

$$\begin{aligned} f_m(x) = (1 - p_W) &\left\{ \tilde{p}_{DW} + f_{m+1}(x) \right\} \\ &+ \min_{z \ge y \ge 0} \left\{ (1 - p_K)^y \tilde{p}_{WW} + f_{m+1}(x - y) \right\} p_W \end{aligned} \tag{9}$$

where z is defined by Eq. (3).

By use of Eqs. (8) and (9) we will be able to determine recursively an optimal assignment of interceptor missiles (0, 1, 2, $\cdots$, or t) to each object in sequence. An optimal number $y_m(x)$ is a function of the stage and the number of interceptor missiles remaining at the stage.

In the preceding derivation of the recursive Eq. (9) we implicitly assumed that all offense missiles would function perfectly in operation, and a formulation for optimizing the assignment of defense interceptor missiles was established on the basis of this assumption. In reality, however, a random number of offense missiles will fail to function successfully, and this fact will certainly affect the mode of assignment. We will briefly consider this offense reliability aspect of the problem in Appendix B.

Finally, as has already been remarked, the preceding formulation does not take into consideration the number of objects called warheads prior to a given stage within the same cloud. The repeated use* of p_W defined by Eq. (1) for all stages is valid only under these circumstances. For the defense to utilize the additional information of the number of objects called warheads prior to a given stage, p_W must be modified from stage to stage depending on this number. We will describe this modified model of the problem in Appendix C.

III. MAX-MINIMIZATION APPROACH TO OPTIMUM DECOY SELECTION

As stated in the introduction, the principle of our approach to optimized decoy selection is first to let the defense choose a discrimination decision threshold that minimizes the expected number of surviving warheads for any given decoy of the offense, and then to let the offense select the decoy specification that produces a maximum among these minimized expected numbers. The expected number is determined from the dynamic programming model for each defense threshold and offense decoy specification considered.

It is assumed that once the offense's decoy specification is definitely set forth, the probability distributions of any discriminable parameter for both warheads and decoys are uniquely determined and known to the defense and to the offense.† Therefore, in a single discriminable parameter case, given the decoy distribution and the defense's selection of a decision threshold, p_{WW} and p_{DW} are determined, as illustrated by Figure 1.

Also every decoy specification will lead to a maximum number, say N, of decoys to be accommodated in an offense missile payload due to the volume and weight constraints.

Once a "triplet" (N, p_{WW}, p_{DW}) is thereby determined by each offense decoy selection and defense threshold selection, an optimal expected number of surviving warheads, say $E(N, p_{WW}, p_{DW})$, will follow from the dynamic single target model by use of Eqs. (8) and (9). Following this, the defense will be able to minimize $E(N, p_{WW}, p_{DW})$ by varying the threshold and so p_{WW} over an appropriate range. Then the offense will associate each such minimized $E(N, p_{WW}, p_{DW})$ with a given decoy, and will select the one that produces a maximum among all these minimized $E(N, p_{WW}, p_{DW})$ values. Making a decoy selection in this way, the offense will be assured that whichever decision threshold the defense may employ, the resulting $E(N, p_{WW}, p_{DW})$ will never be smaller than the number associated

*It can be seen that the case considered is analogous to probability theory's urn drawing with replacement. The more realistic situation's analog is drawing without replacement.

†In practice, of course, each estimates the statistics of these distributions. Mismatches in such estimates, possibly of great significance, are not considered here.

with the selected specification. The specification is consequently optimal in a max-min sense.

IV. NUMERICAL EXAMPLE

The example presented herein illustrates the application of the decoy optimization procedure to the following situation. The offense attacks a target with three missiles, each missile containing one warhead and N decoys, with the variable N to be determined by the optimization procedure under the constraint that the total weight of each decoy package cannot exceed 150 pounds.* Each missile is assumed to be perfectly reliable; that is, there is a probability of one that each offensive cloud (consisting of one warhead and N decoys) will present a threat to the target and thus cause damage to the target if not successfully killed by the defense. The defense is assumed to have 18 interceptor missiles, each with a kill probability against a warhead of .80. The restriction is made that not more than three interceptor missiles are to be fired per target.

The defense's discrimination is to be based upon two parameters, ballistic coefficient β and radar cross section σ. An object must "pass" both tests in order to be classified as a warhead.

As was discussed in Section II, an object is classified as a warhead if the measured discriminable parameters (σ and β in the example) exceed their respective threshold values, say τ_1 and τ_2. The warhead distribution for each parameter is fixed for any given attack situation and, for a particular decoy design, the decoy distributions are also fixed. The corresponding values of p_{DW} and p_{WW} are then determined as discussed in Appendix A (also refer to Figure 1).

For this example, the β distribution for the warhead is assumed normal, with a mean of 1,000 lbs/ft^2 and a standard deviation of 200 lbs/ft^2. The warhead σ distribution is also assumed normal, with a mean of 1 ft^2 and a standard deviation of .10 ft^2. The decoy distributions are also taken to be normal, with the standard deviation of the σ distribution equal to 10 percent of the mean and the β standard deviation to be 20 percent of the mean. The relation between unit decoy weight w and the decoy characteristics as measured by the two parameters σ and β is the illustrative scaling law,†

$$w = .025\,\bar{\sigma}^{1/2}\bar{\beta} \tag{10}$$

An ideal decoy, that is, one whose $\bar{\sigma}$ and $\bar{\beta}$ match the warhead's $\bar{\sigma}$ and $\bar{\beta}$, then weighs 25 pounds. Thus a total of 6 ideal decoys can be carried per missile or, of course, a greater number of less-than-ideal decoys.

In order to determine the optimal decoy package specifications (N, $\bar{\beta}$, $\bar{\sigma}$; w) under the assumptions discussed above, a computer program was

*A volume constraint is not considered here.

†The barred symbols represent the mean values of the parameters involved.

developed which computes, on the basis of the model discussed in this paper, the expected number of surviving warheads for each pair of values of the decoy $\bar{\sigma}$ and $\bar{\beta}$ parameters. In doing so, two defense minimizations are performed. First, for a given decoy configuration, as specified by $\bar{\sigma}$ and $\bar{\beta}$, and a given pair of threshold values τ_1 and τ_2, the minimum expected number of surviving warheads E which the defense can attain by optimal interceptor missile allocation is determined by application of the dynamic programming model discussed in Section II. Then, with the decoy $\bar{\sigma}$ and $\bar{\beta}$ still fixed, the values of τ_1 and τ_2 that minimize the resulting E are determined. Each set of fixed values of τ_1, τ_2, decoy $\bar{\sigma}$, and decoy $\bar{\beta}$ result in a particular triplet (N, p_{WW}, p_{DW}), as explained in Section III. Finally, the offense maximizes over these minimum values of E by varying the decoy design parameters. This, then, results in the max-min solution of the two-sided game referred to earlier.

Table 1 presents the resulting values of E and N for various pairs of values of decoy $\bar{\sigma}$ and $\bar{\beta}$. These E values represent the defense-attained minimums in each case; that is, each represents the best the defense is able to do by optimal interceptor missile allocation and optimal selection of the thresholds.

Table 1. Table of Values of E and N for Decoy Packages With Parameters $\bar{\sigma}$, $\bar{\beta}$

	$\bar{\beta}$ (lbs/ft^2)											
$\bar{\sigma}$	700		800		850		900		950		1,000	
(ft^2)	E	N	E	N	E	N	E	N	E	N	E	N
.70	.20	10	.20	9	.19	8	.19	8	.19	8	.19	7
.80	.50	10	.45	8	.45	8	.43	7	.43	7	.43	7
.85	.64	9	.70	8	.70	8	.64	7	.64	7	.64	7
.90	.70	9	.88	8	.82	7	.82	7	.82	7	.71	6
.95	.70	9	.92	8	.94	7	1.01	7	.85	6	.85	6
1.00	.71	9	.92	8	.94	7	1.03	7	.91	6	.94	6

It can be readily seen that the best decoy from the offense's point of view is one with $\bar{\sigma}$ = 1.00 ft^2 and $\bar{\beta}$ = 900 lbs/ft^2. A total of 7 of these decoys, each of weight 20.7 pounds, can be carried per missile. The resulting expected number of surviving warheads is 1.03 out of a total of 3. The values of E are relatively low because there are 18 interceptor missiles assigned to a total of 24 targets, allowing almost 1 per target on the average. It can be seen from Table 1 that there is a significant difference between the effectiveness of the optimal decoy package and the ideal decoy package, indicating the advantage of utilizing the optimization procedure in determining the specifications of a decoy package.

A table representing the optimal defense allocation of interceptor missiles per target (with a limit of three interceptor missiles per target) for each stage (of which there are $P(N + K) = 24$) for an attack employing the optimal decoy package, is presented in Figure 2. The horizontal axis indexes the stage of the process at which an assignment decision is to be made of

developed which computes, on the basis of the model discussed in this paper, the expected number of surviving warheads for each pair of values of the decoy $\bar{\sigma}$ and $\bar{\beta}$ parameters. In doing so, two defense minimizations are performed. First, for a given decoy configuration, as specified by $\bar{\sigma}$ and $\bar{\beta}$, and a given pair of threshold values τ_1 and τ_2, the minimum expected number of surviving warheads E which the defense can attain by optimal interceptor missile allocation is determined by application of the dynamic programming model discussed in Section II. Then, with the decoy $\bar{\sigma}$ and $\bar{\beta}$ still fixed, the values of τ_1 and τ_2 that minimize the resulting E are determined. Each set of fixed values of τ_1, τ_2, decoy $\bar{\sigma}$, and decoy $\bar{\beta}$ result in a particular triplet (N, p_{WW}, p_{DW}), as explained in Section III. Finally, the offense maximizes over these minimum values of E by varying the decoy design parameters. This, then, results in the max-min solution of the two-sided game referred to earlier.

Table 1 presents the resulting values of E and N for various pairs of values of decoy $\bar{\sigma}$ and $\bar{\beta}$. These E values represent the defense-attained minimums in each case; that is, each represents the best the defense is able to do by optimal interceptor missile allocation and optimal selection of the thresholds.

Table 1. Table of Values of E and N for Decoy Packages With Parameters $\bar{\sigma}$, $\bar{\beta}$

$\bar{\sigma}$	$\bar{\beta}$ (lbs/ft^2)											
	700		800		850		900		950		1,000	
(ft^2)	E	N	E	N	E	N	E	N	E	N	E	N
.70	.20	10	.20	9	.19	8	.19	8	.19	8	.19	7
.80	.50	10	.45	8	.45	8	.43	7	.43	7	.43	7
.85	.64	9	.70	8	.70	8	.64	7	.64	7	.64	7
.90	.70	9	.88	8	.82	7	.82	7	.82	7	.71	6
.95	.70	9	.92	8	.94	7	1.01	7	.85	6	.85	6
1.00	.71	9	.92	8	.94	7	1.03	7	.91	6	.94	6

It can be readily seen that the best decoy from the offense's point of view is one with $\bar{\sigma}$ = 1.00 ft^2 and $\bar{\beta}$ = 900 lbs/ft^2. A total of 7 of these decoys, each of weight 20.7 pounds, can be carried per missile. The resulting expected number of surviving warheads is 1.03 out of a total of 3. The values of E are relatively low because there are 18 interceptor missiles assigned to a total of 24 targets, allowing almost 1 per target on the average. It can be seen from Table 1 that there is a significant difference between the effectiveness of the optimal decoy package and the ideal decoy package, indicating the advantage of utilizing the optimization procedure in determining the specifications of a decoy package.

A table representing the optimal defense allocation of interceptor missiles per target (with a limit of three interceptor missiles per target) for each stage (of which there are $P(N + K) = 24$) for an attack employing the optimal decoy package, is presented in Figure 2. The horizontal axis indexes the stage of the process at which an assignment decision is to be made of

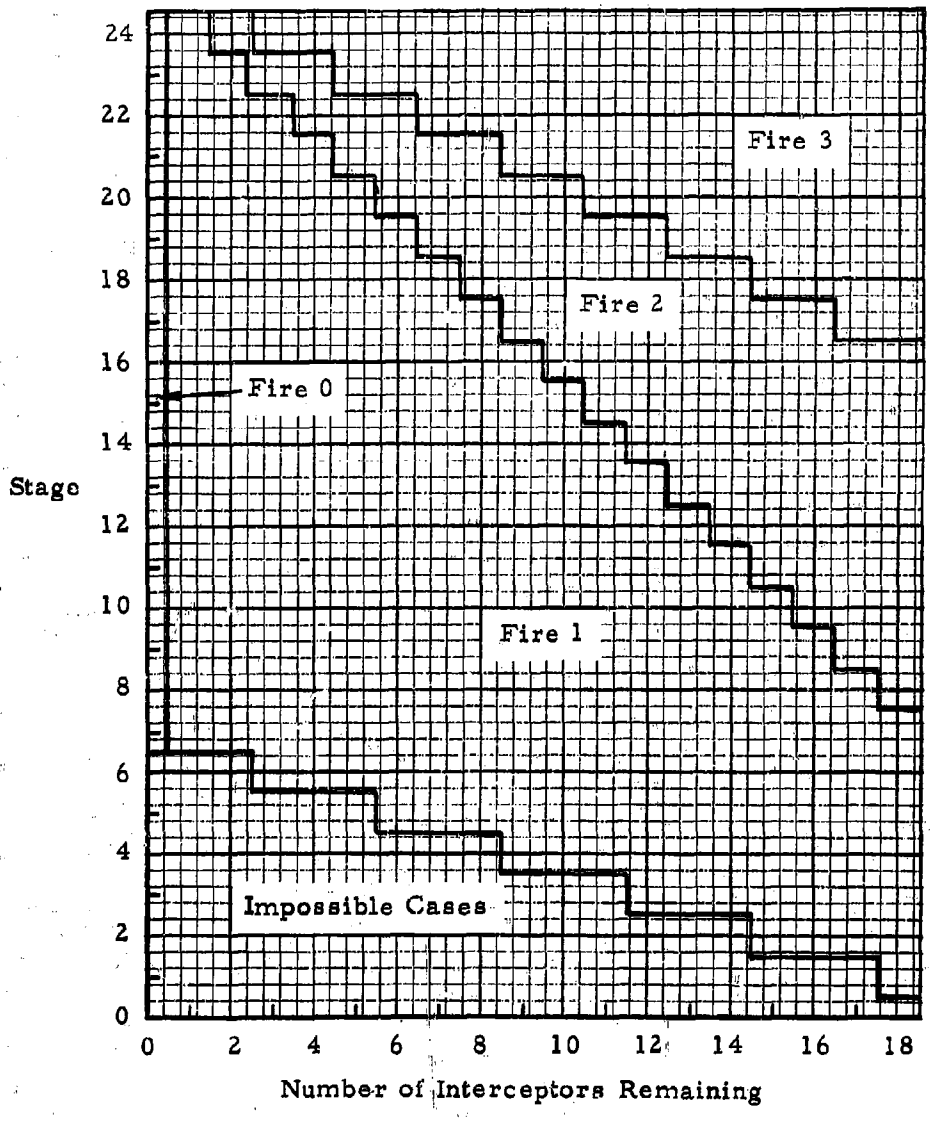


Fig. 2. Optimal Defense Allocation Doctrine Against Optimized Decoy Package

zero, one, two, or three interceptors to an apparent warhead. The vertical axis indexes the number of interceptors still available for use. The assignments fall into regions corresponding to firing 0, 1, 2, or 3 interceptors, as indicated. Impossible cases, in which so few interceptors cannot remain after so few stages due to the 3-shot-per-target limitation assumed, also fall into a region as shown.

V. SUMMARY

A model has been presented for the terminal battle between a sequence of attacking ballistic missiles and a Nike-Zeus-type defense. The model enables selection of optimal reentry decoy package specifications against a defense optimized against this package in its interceptor allocation doctrine and discrimination decision rules. This selection employs the model to evaluate a measure of effectiveness, the expected number of warheads in the attack that survive the defense, for each offense and defense choice of parameters, and then solves the (two-sided, perfect information) game for a max-min decoy package specification. A numerical example indicates that optimal (more, less-than-perfect) decoys can provide significantly greater effectiveness than (fewer) ideal ones, for given payload constraints.

APPENDIX A
MULTIPLE DISCRIMINABLE PARAMETERS

In the main body of this paper, only a single discrimination test was considered for simplicity in exposition. Generally, the defense's discrimination test may be based on any number of parameters. While we will describe here a procedure that is explicitly applicable to a case involving only two parameters, the manner in which some extensions can be made will be readily seen.

If two independent defense discrimination tests are applied to an object, and an "apparent" warhead is defined to be an object that passes both of these tests, then we have

$$p_{WW} = p_{WW_1} p_{WW_2} \tag{A-1}$$

and

$$p_{DW} = p_{DW_1} p_{DW_2} \tag{A-2}$$

where p_{WW_i} is the probability that a warhead will pass the test i, and p_{DW_i} is the probability that a decoy will pass the test i. The individual test acceptance probabilities, p_{WW_i} and p_{DW_i}, are determined, as in the single discriminable parameter case, by the defense's selection of decision thresholds, τ_i, and the offense's choice of the distribution of decoy parameters or decoy specification, with a consequent determination of N. Thus, for a given pair of thresholds (τ_1, τ_2) and a given decoy specification, there corresponds a triplet (N, p_{WW}, p_{DW}), as in the single discriminable parameter case, which will lead to an optimal expected number of surviving

warheads, $E(N, p_{WW}, p_{DW})$, by use of Eqs. (8) and (9) of the dynamic single target model.

An important observation to be made here, however, is that there may exist a number of different pairs (τ_1, τ_2) that produce an identical p_{WW}, and consequently, there may exist a number of different p_{DW}'s corresponding to a given p_{WW}. While this fact does not affect our max-minimization approach in its essentiality, it does increase the necessary amount of computation. (This will be so even more in a case involving more than two parameters.) Fortunately there exists a means to circumvent this difficulty which is based on the following provable fact: $E(N, p_{WW}, p_{DW})$ is monotonically increasing in p_{DW} for fixed N and p_{WW}; that is, we have, for $p'_{DW} > p''_{DW}$,

$$E(N, p_{WW}, p'_{DW}) \geq E(N, p_{WW}, p''_{DW}) \quad . \tag{A-3}$$

Because of this monotonicity of $E(N, p_{WW}, p_{DW})$ for fixed N and p_{WW}, we are assured to minimize $E(N, p_{WW}, p_{DW})$ for given N and p_{WW} as required, if we select a p_{DW} that is a minimum among all p_{DW}'s corresponding to all possible (τ_1, τ_2) pairs that produce a given p_{WW}. It is comforting to notice that this minimization of p_{DW} can be accomplished outside of the computation of $E(N, p_{WW}, p_{DW})$, and this will therefore entail a considerable savings in computation.

Now that we have a procedure to obtain a specific p_{DW} for a given p_{WW}, the decoy selection problem will be treated in the identical manner as in the single discriminable parameter case. The defense will first vary p_{WW} over an appropriate range of values to obtain a minimum $E(N, p_{WW}, p_{DW})$ for each given decoy possibility, and the offense then will select the decoy specification that will maximize these minimized $E(N, p_{WW}, p_{DW})$ values.

APPENDIX B

MODIFICATION TO INCLUDE OFFENSE MISSILE UNRELIABILITY

In this appendix, we will explicitly take into account the unreliability factor of an offense missile in applying dynamic programming to the problem of assignment of defense interceptor missiles. Let us designate by p_R the probability that an offense missile will successfully arrive at the vicinity of a target and eject the warheads and decoys so that the cloud of the missile will become a target of the defense interception. Due to this probability, the total number of clouds the defense will be facing is a random variable, even if the total number of offense missiles to be fired at the target is definitely known. The probability that precisely k offense missiles out of P will be successfully transformed into target clouds is given by

$$\binom{P}{k} p_R^k (1 - p_R)^{P-k} \qquad 0 \leq k \leq P \quad . \tag{B-1}$$

Suppose now that we are at stage nL, during the battle, $1 \leq n < P$; that is, we are faced with the last object of cloud n. Since the total number of

clouds the defense will encounter in a present problem is a random variable, we cannot be certain that we will have to deal with cloud $n+1$; that is, that the process will continue to stage $nL+1$. Therefore, the recursive Eq. (9) must be suitably modified for $m = nL$, $1 \le n < P$. The modification, however, is very slight. The only thing we have to do is to determine explicitly the aforementioned uncertainty of the process in terms of probability p_R and discount the contribution to the expected number of surviving warheads from stage $nL+1$ onwards on the basis of this uncertainty. Specifically, if we designate by p_n the probability that the process will continue to stage $nL+1$ given that it has reached stage nL, then the modified recursive equation will read

$$f_m(x) = (1 - \tilde{p}_W)\left\{p_{DW} + p_n f_{m+1}(x)\right\} + \min_{z \ge y \ge 0}\left\{(1 - \tilde{p}_K)^y p_{WW} + p_n f_{m+1}(x-y)\right\} p_W \qquad \text{(B-2)}$$

for $m = nL$, $1 \le n < P$.

Now let us determine the probabilities p_n. As stated before, p_n is the probability that cloud $n+1$ will arrive when clouds 1 through n have already arrived. If we identify an arrival of a cloud as a success in a Bernoulli trial with probability p_R of success, then any set of successive clouds the defense is going to encounter may be considered as a result of a sequence of P identical Bernoulli trials with probability p_R of success. Thus, the event that the defense has already encountered n clouds corresponds to a sequence of P outcomes in which there are at least n successes.

The probability of occurrence of a sequence of P outcomes with at least n successes is given by

$$\sum_{k=n}^{P} \binom{P}{k} p_R^k (1 - p_R)^{P-k} .$$

The conditional probability p_n is the ratio of the probabilities of occurrence of sequences with $n+1$ and N successes, respectively. Therefore,

$$p_n = \left[\sum_{k=n+1}^{P} \binom{P}{k} p_R^k (1 - p_R)^{P-k}\right] \Bigg/ \left[\sum_{k=n}^{P} \binom{P}{k} p_R^k (1 - p_R)^{P-k}\right] . \qquad \text{(B-3)}$$

APPENDIX C

MODIFICATION OF ASSIGNMENT PROCESS TO TAKE INTO ACCOUNT NUMBER OF APPARENT WARHEADS IN PRIOR STAGES OF A CLOUD

In this appendix we will explicitly take into consideration the number of objects called warheads by the defense prior to a given stage within the same cloud. Let us suppose that we are faced with the jth object of the ith cloud,

and that u objects among the preceding j - 1 objects within the cloud i were called warheads.

We define the following probabilities:

Probability of having had k (true) warheads among the preceding j - 1 objects:

$$p_j(k) = \binom{K}{k} \binom{N}{j-1-k} \Big/ \binom{L}{j-1} \tag{C-1}$$

Probability of calling u objects (apparent) warheads among the preceding j - 1 objects, given that there were k warheads among the preceding j - 1 objects:

$$p_j(u;k) = \sum_i \binom{k}{i} \binom{j-1-k}{u-i} (p_{WW})^i (1-p_{WW})^{k-i} (p_{DW})^{u-i} (1-p_{DW})^{j-1-k-u+i} \tag{C-2}$$

where $0 \le i \le \min(k, u)$.

Probability of having had k warheads among the preceding j - 1 objects, given that there were u objects called warheads among the preceding j - 1 objects:

$$p_j(k;u) = p_j(u;k)\, p_j(k) / \sum_k p_j(u;k)\, p_j(k) \tag{C-3}$$

where $\max(0, j-1-N) \le k \le \min(K, j-1)$.

Probability of calling the jth object a warhead when there are K - k warheads among the remaining L - j + 1 objects; that is, when there were k warheads among the preceding j - 1 objects:

$$p_W(j;k) = \frac{K-k}{L-j+1}\, p_{WW} + \frac{N+k-j+1}{L-j+1}\, p_{DW} \tag{C-4}$$

Probability that the jth object is called a warhead, given that there were u objects called warheads among the preceding j - 1 objects:

$$p_W(j;u) = \sum_k p_W(j;k)\, p_j(k;u) \tag{C-5}$$

where $\max(0, j-1-N) \le k \le \min(K, j-1)$

Probability that the jth object is called a warhead when it is in fact a warhead, given that there were u objects called warheads among the preceding j - 1 objects:

$$p_{WW}(j;u) = \sum_k \frac{K-k}{L-j+1}\, p_{WW}\, p_j(k;u) \tag{C-6}$$

where $\max(0,\ j-1-N) \le k \le \min(K,\ j-1)$.

Probability that the jth object is a true warhead, given that it was called a warhead, and that there were u objects called warheads among the preceding j-1 objects:

$$\tilde{p}_{WW}(j;u) = p_{WW}(j;u)/p_W(j;u) \quad . \tag{C-7}$$

Probability that the jth object is called a decoy when it is in fact a warhead, given that there were u objects called warheads among the preceding j-1 objects:

$$p_{WD}(j;u) = \sum_k \frac{K-k}{L-j+1}\,(1-p_{WW})\,p_j(k;u) \tag{C-8}$$

where $\max(0,\ j-1-N) \le k \le \min(K,\ j-1)$.

Probability that the jth object is a true warhead, given that it was called a decoy, and that there were u objects called warheads among the preceding j-1 objects:

$$\tilde{p}_{DW}(j;u) = p_{WD}(j;u)/[1-p_W(j;u)] \quad . \tag{C-9}$$

If we now designate by $f_{iL+j}(x,u)$ the minimum expected number of surviving warheads from the jth stage of the ith cloud onwards, when there are x interceptor missiles remaining at the stage and there were u objects called warheads among the preceding j-1 objects within the cloud i, then we obtain, by use of Eqs. (C-5), (C-7), and (C-9), and by reasoning identical to that of Section II, the following recursive equation analogous to Eq. (9):

$$\begin{aligned} f_{iL+j}(x,u) = {} & [1-p_W(j;u)]\,[\tilde{p}_{DW}(j;u) + f_{iL+j+1}(x,u)] \\ & + \min_{z\ge y\ge 0} [(1-p_K)^y\,\tilde{p}_{WW}(j;u) + f_{iL+j-1}(x-y,u+1)]\,p_W(j;u) \end{aligned} \tag{C-10}$$

where z is defined by Eq. (3). An optimal number of assigned interceptors $y_{iL+j}(x,u)$ is now a function of the stage, the number of interceptor missiles remaining at the stage, and the number of objects called warheads prior to the stage within the same cloud.

EXPERIMENTAL STUDY OF THE EFFECTS OF SURFACE IRREGULARITIES ON THE HYPERSONIC AERODYNAMIC HEAT TRANSFER TO A FLAT PLATE

Paul G. Kafka
Loren H. Anderson

The Boeing Company
Aerospace Division
Seattle 24, Washington

ABSTRACT

The effects of surface irregularities upon aerodynamic heat transfer in hypersonic flow have been investigated.

Heat transfer coefficients and pressures have been measured in the presence of notches, forward and backward facing steps, and sine wave type corrugations normal to the flow.

All these irregularities caused separation and reattachment of the flow. The measured quantities are ratioed both to laminar and turbulent flat plate values.

The effects of Mach number, Reynolds number, and angle of attack were also investigated. All irregularities were found to cause localized heating peaks, some with magnitudes over twice turbulent flat plate values.

The testing program was done under USAF Contract AF33(600)-41517.

EXPERIMENTAL STUDY OF THE EFFECTS OF SURFACE IRREGULARITIES ON THE HYPERSONIC AERODYNAMIC HEAT TRANSFER TO A FLAT PLATE

Paul G. Kafka and Loren H. Anderson
The Boeing Company
Aerospace Division
Seattle 24, Washington

INTRODUCTION

In manufacturing large space vehicles with complicated shapes it is very difficult, if not impossible, to obtain perfectly smooth exterior surfaces.

Expansion gaps between skin panels, counter-sunk rivets, skin buckling due to temperature gradients and other causes will confront the flow with small irregularities in surface geometry. These irregularities will basically be of three kinds: grooves, steps (forward and backward facing) and waves; they can all be expected to induce local boundary layer separation with downstream reattachment and localized hot spots.

Since no satisfactory theories for reattachment heating exist at present, the authors' company decided to run an extensive testing program in order to get a better understanding of these phenomena.

The models consisted basically of flat plates with removable inserts, so that heating rates at a given station both with and without the irregularity could be compared. Mach numbers ranged from 6 to 9.6 and Reynolds numbers from 5×10^5 to 18×10^6 per ft.

MODELS

Three basic models were used. Model I (Fig. 1) is a blunt flat plate of square planform with a hollow space for inserts in the aft section. Four inserts were provided:

Insert Ia provides the basic flat plate configuration.

Insert Ib creates a forward facing step followed by a backward facing one. Step height was .05 in., step length 1 in.

Insert Ic was a full sine wave dip with .1 in. amplitude and 1.5 in. wave length.

Insert Id contained two notches .05 in. and .1 in. deep, respectively, and width equal to depth.

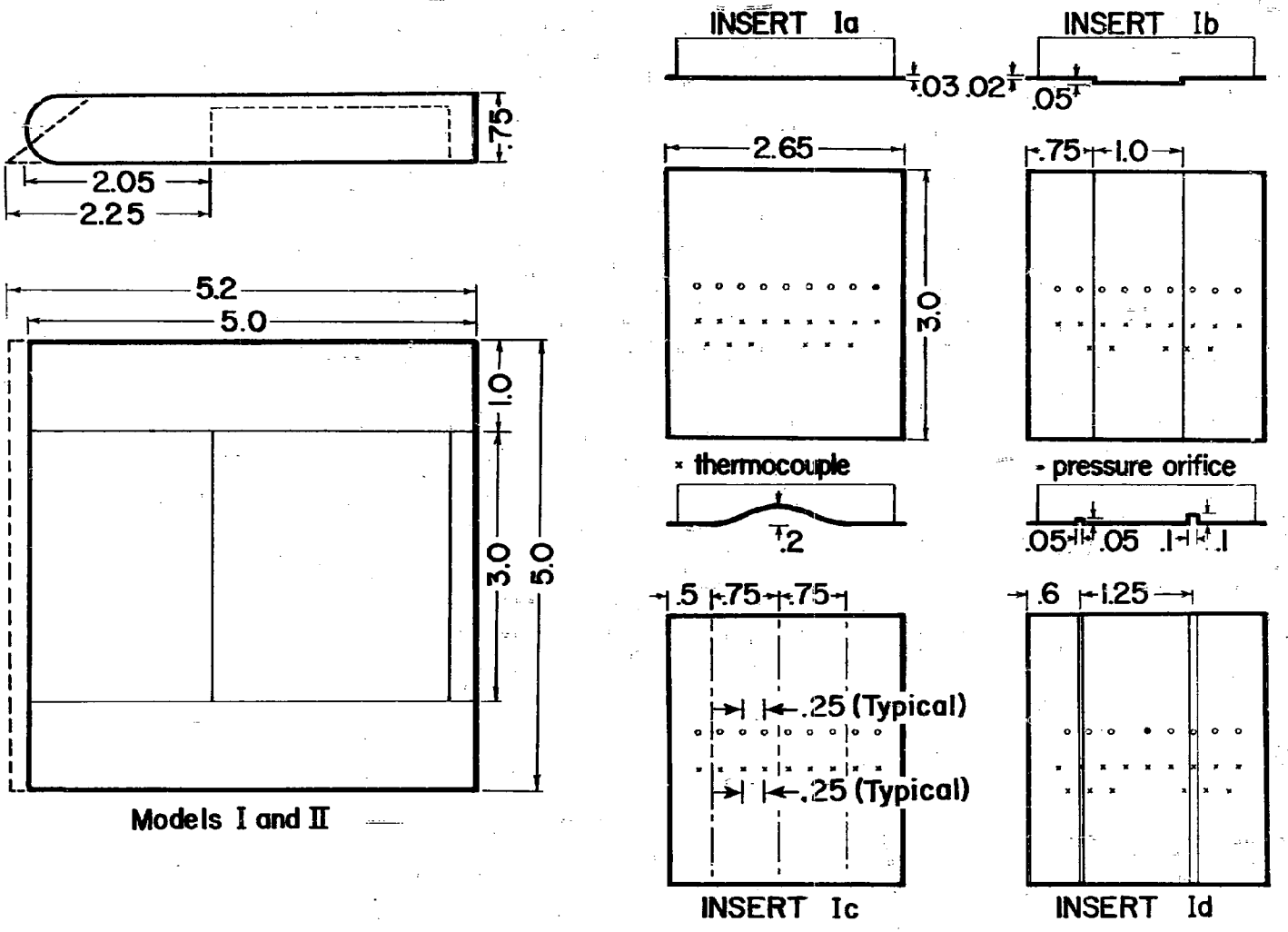


Fig. 1. Models I and II with Removable Leading Edges and Roughness Inserts

A photograph of Model I with its inserts is shown in Fig. 2.

Model II (Fig. 1) was identical to I except that the blunt nose had been replaced by a sharp nose (37° wedge). Inserts were identical to I.

Model III (Fig. 3) was similar to I except for size, being twice as large. Inserts IIIa, IIIb, IIId are similar to Ia, Ib, Id, the depth of the irregularities being the same. Insert IIIc consisted of two sine wave dips in series, each one of the same amplitude and wave length as in I

All models were made of 321 stainless steel with wall thickness of .02 in. and .03 in. Only the inserts were instrumented with thermocouples and pressure taps, as shown in the figures.

EXPERIMENTAL FACILITIES

Models I and II were tested in the Boeing Hypersonic Wind Tunnel (BHWT) and the NASA Langley Research Center (LRC) 11 in. tunnel.

Model III was tested in the Arnold Center "B" tunnel. This model suffered considerably from thermal warping and sting effects, and only occasional reference to this test will be made here. For a detailed report see (1).

Test conditions were:

Boeing Hypersonic Wind Tunnel:

Mach: 6
Re/ft: 13 and 18 x 10^6
Total Conditions were:
For R_e = 13 x 10^6/ft
T_t = 1010°R
P_t = 900 psi
For R_e = 18 x 10^6/ft
T_t = 900°R
P_t = 1100 psi

The Boeing Hypersonic Wind Tunnel is a blow down type tunnel, with a test period of 30 to 60 seconds.

The model is pre-cooled and then injected into the tunnel. Model skin temperatures are measured by means of Chromel-Alumel thermocouples spotwelded to the inner surface of the skin. The transient temperature method was used to obtain heat transfer coefficients. A computer program handled the task of converting the thermocouple output into temperatures, smoothing the data to obtain initial slopes of temperature-time plots, and calculating heat flows and adiabatic wall temperatures. Skin conductivity corrections proved to be negligible.

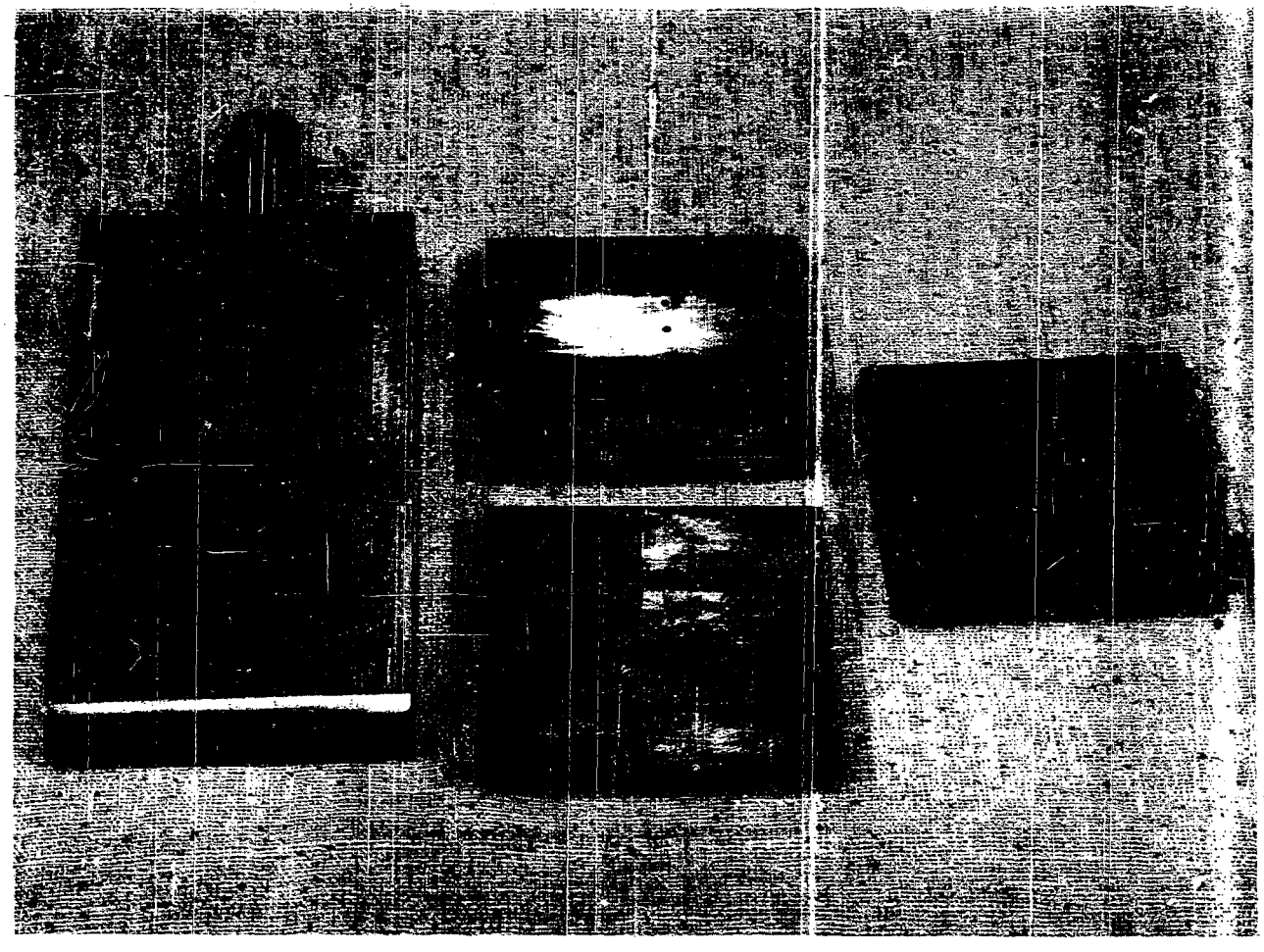


Fig. 2. Basic Model with Inserts: (Ia) Flat Plate (Installed), (Ib) Steps, (Ic) Sine Wave, and (Id) Notches

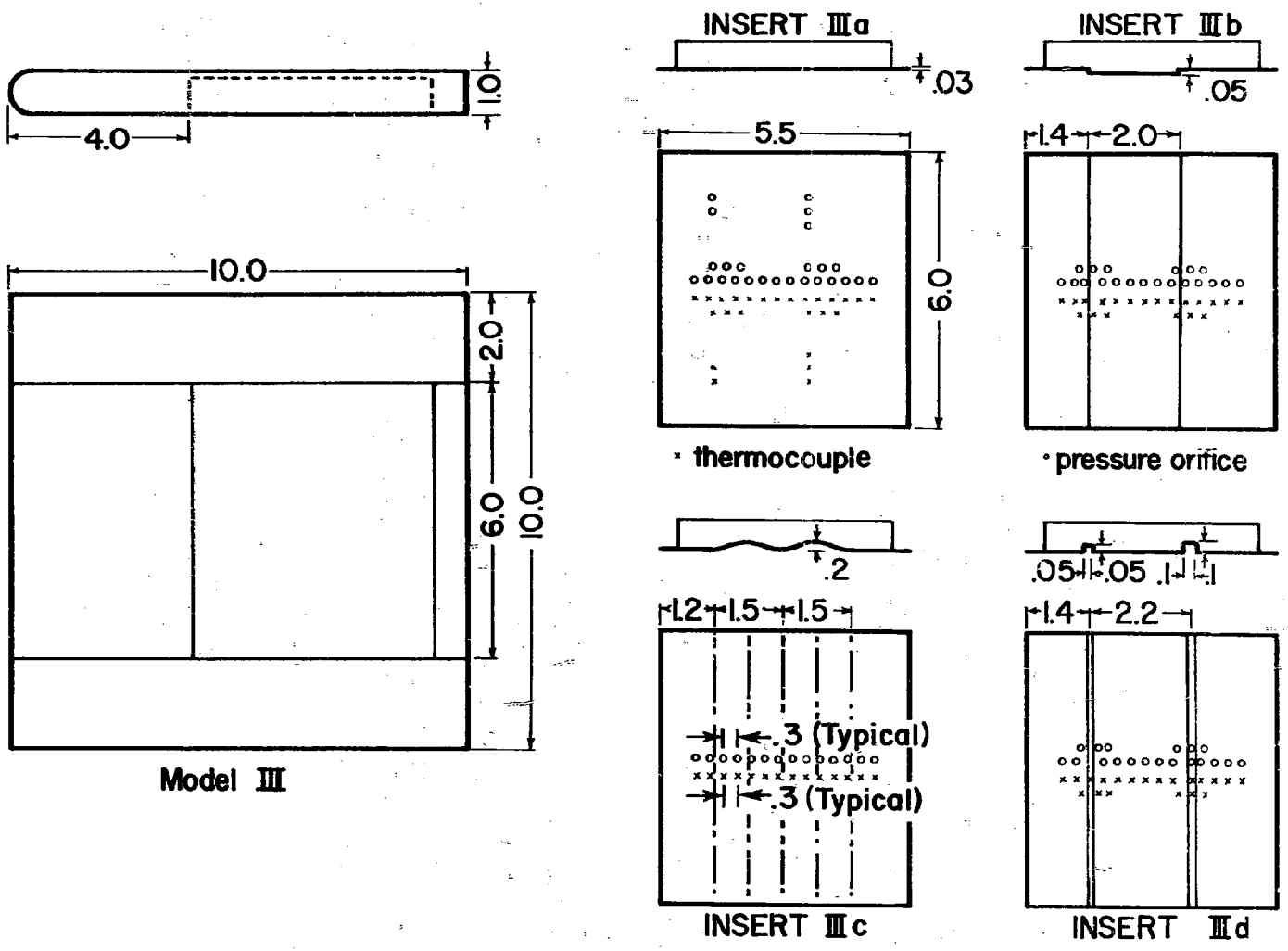


Fig. 3. Model III and Roughness Inserts

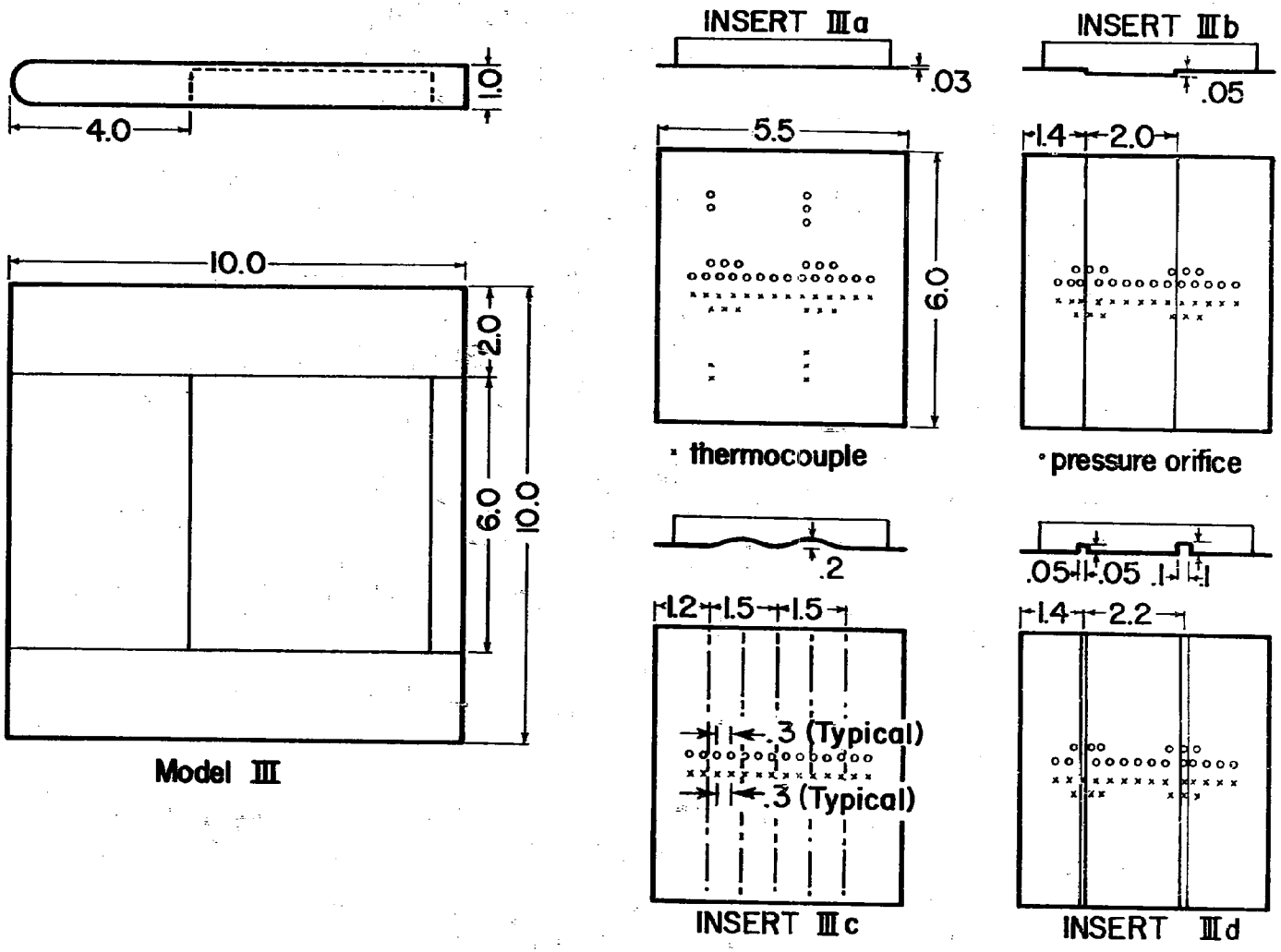


Fig. 3. Model III and Roughness Inserts

Langley 11 in. Tunnel

Mach: 9.6
Re/ft: $.8 \times 10^6$ and 1.2×10^6
Total Conditions were:
For $R_e = .8 \times 10^6$ per ft
$T_t = 1600°R$
$P_t = 380$ psi
For $R_e = 1.2 \times 10^6$ per ft
$T_t = 1600°R$
$P_t = 675$ psi

The Langley 11 in. Hypersonic Tunnel is a blow down type facility which utilizes the bypass technique. The model is cooled to a uniform temperature, the flow is started while bypassing the test section and switched to the latter when fully established.

Instrumentation was identical to the BHWT models. All data are reduced by hand. Conduction corrections were considerable and were evaluated by comparing the slopes of the time-temperature records at adjacent thermocouples in the flow direction. A picture of Model I installed in this tunnel is shown in Fig. 4. The model was equipped with end plates in this test, in order to assure two-dimensional flow, although little change was observed.

Arnold Center "B" Tunnel

Mach: 8
Re/ft: $.5 \times 10^6$ and 1.5×10^6
Total Conditions were:
For $R_e = .5 \times 10^6$ per ft
$T_t = 1335°R$
$P_t = 115$ psi
For $R_e = 1.5 \times 10^6$ per ft
$T_t = 1335°R$
$P_t = 350$ psi

The AC "B" Tunnel is a continuous tunnel which utilizes a "cooling shoe" to precool the model. To obtain a transient temperature condition, the cooling shoes pop open and the thermocouples give a digital record of temperatures at 1/20 second intervals for 10 sec. The procedure for converting thermocouple readings into heat transfer coefficients was analogous to the one used at BHWT. Conductivity corrections were found negligible.

Tests in the BHWT were run both with and without a tripping wire. The latter had a diameter of .0126 in. and was located 1.5 in. ahead of the insert, about .03 in. above the plate.

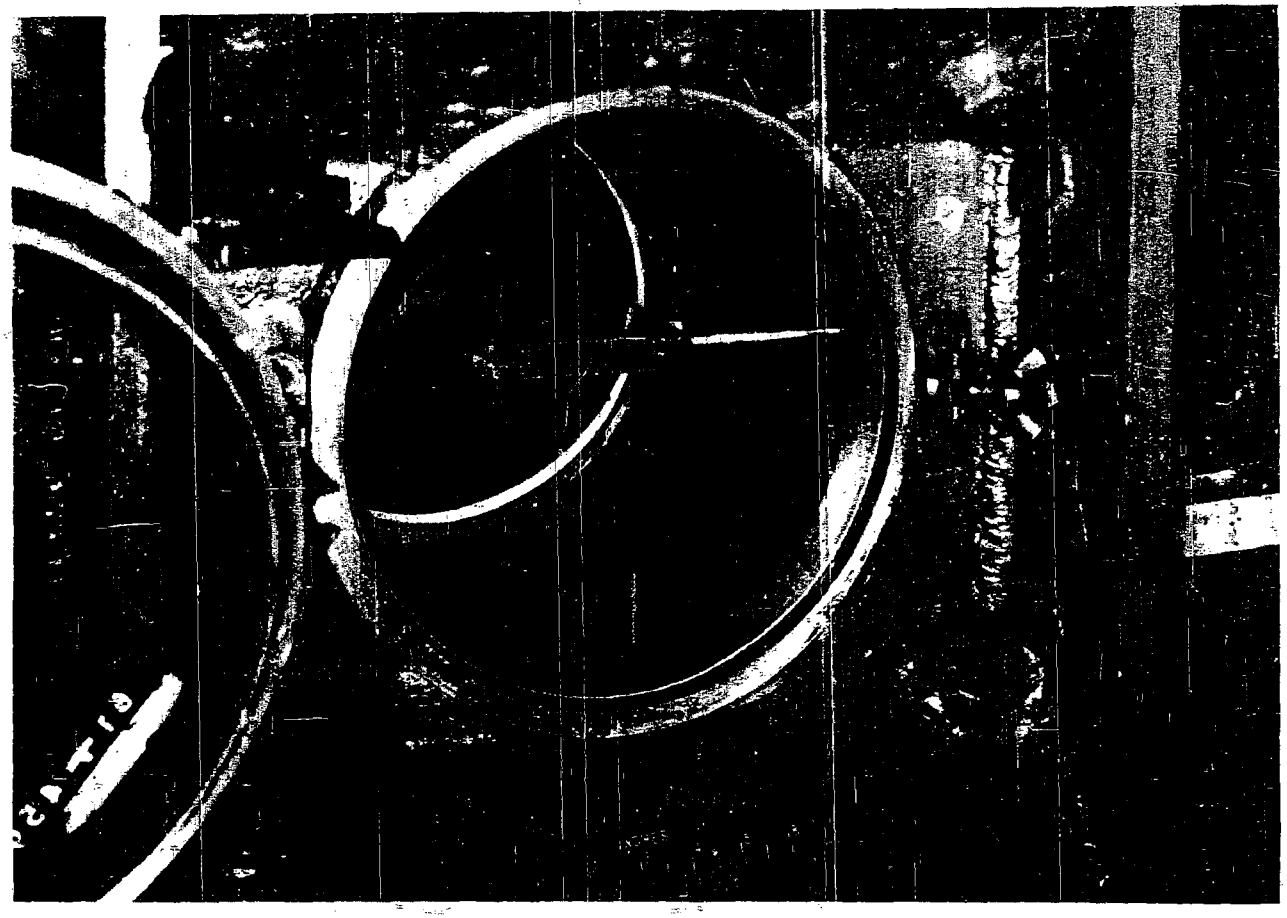


Fig. 4. Model I Installed in Langley Research Center, 11-Inch Hypersonic Facility

Plots of theoretical boundary layer displacement thicknesses vs. angle of attack at a station just upstream of the forward facing step are shown in Fig. 5 for all test conditions. They are based upon compressible flat plate theories, namely the Monaghan formula for laminar flow and the Prandtl formula with reference temperature, for turbulent flow.

RESULTS

(1) BASIC FLAT PLATES

Fig. 6 shows the basic blunt flat plate (Model I) heat transfer at α = 0, presented in terms of the parameter $N_{ST_\infty}\sqrt{Re_{\infty,t}}$ where $N_{ST_\infty}=h/\rho_\infty u_\infty c_{p_\infty}$ and $Re_{\infty,t}=u_\infty \rho_\infty t/\mu_\infty$. h is the local heat transfer coefficient, t is the plate thickness, u_∞, ρ_∞, c_{p_∞}, μ_∞ are free stream velocity, density, specific heat, and viscosity, respectively.

Good agreement with turbulent theory was obtained and the tripped case shows that the trip was highly effective in producing a turbulent boundary layer. The agreement was less satisfactory for the laminar cases.

Fig. 7 shows the same model at α = 20°. It can be seen that at this angle of attack the BHWT boundary layer was turbulent or transitional, even without trip.

Fig. 8 shows the heat transfer to the sharp flat plate (Model II) at α = 0 and α = 20°. The flow remained laminar throughout.

Theoretical curves are based upon the Blasius equation for laminar flow and the Colburn equation for turbulent flow, both with temperature dependent parameters evaluated at the Eckert reference temperature.

A photograph of Model I is shown in Figs. 9 and 10, untripped and tripped, respectively.

(2) STEPS

Heat flow is presented in terms of the parameter $N_{ST_\infty}\sqrt{Re_{\infty,t}}$.

Figs. 11, 12, 13 show the ratio of this parameter for the step to its value for the flat plate, plotted vs. station.

Fig. 11 shows a comparison at α = 0 of the LRC tests (low Reynolds number) with the BHWT tests (high Reynolds number), the latter both tripped and untripped.

An examination of the corresponding flat plate runs, shown in Fig. 6, shows that for two of them (the untripped ones) the basic flat plate flow was laminar or slightly transitional, while for the tripped run the basic flat plate was fully turbulent. The ratios plotted in Fig. 11 indicate that the step induced transition in the separated flow, with a return to laminar downstream of the step and a new transition at reattachment after the backward facing step.

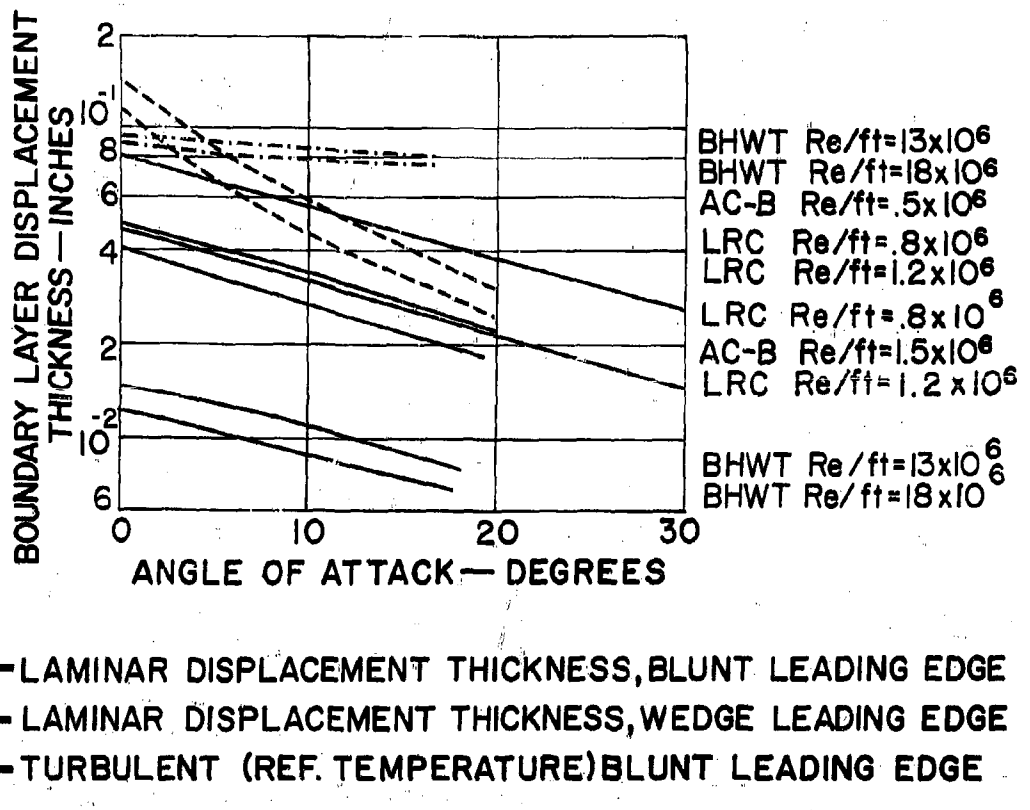


Fig. 5. Flat Plate Displacement Boundary Layer Thicknesses at Step Location for Various Test Conditions

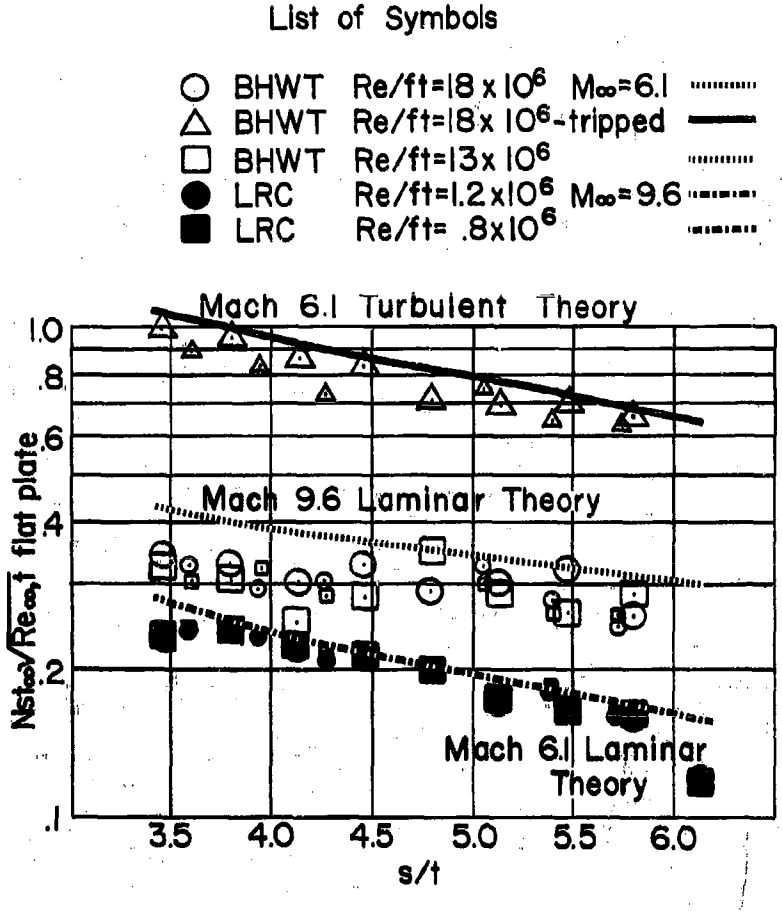


Fig. 6. Blunt Leading Edge Flat Plate at Angle of Attack = 0°, with Variable Reynolds Number

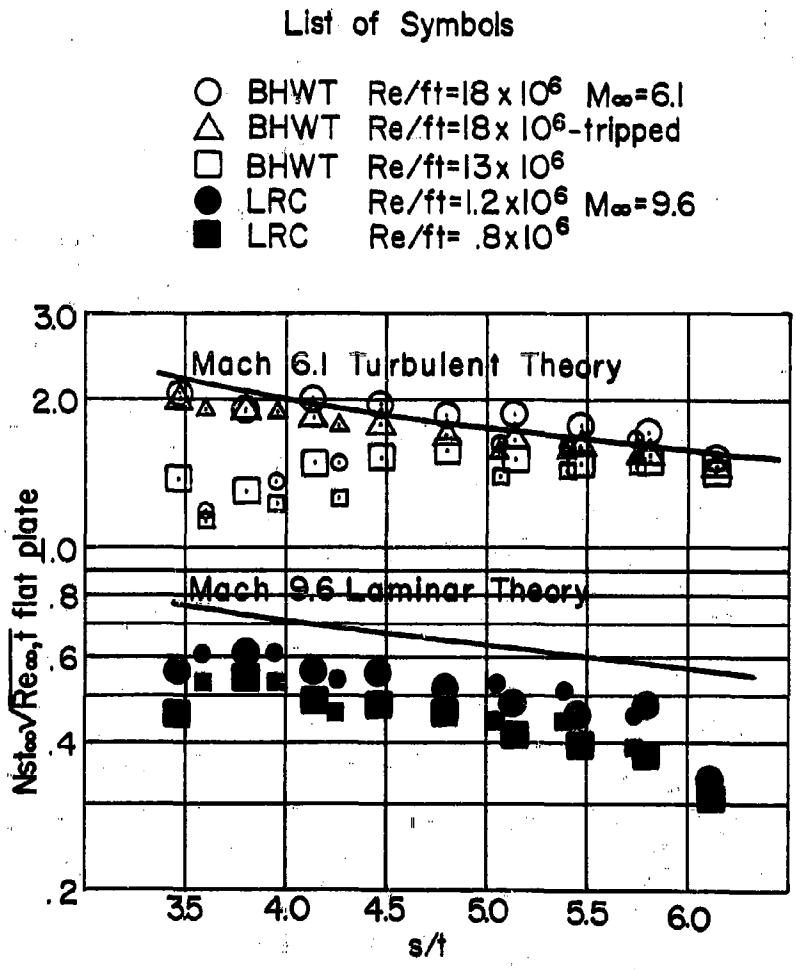


Fig. 7. Blunt Leading Edge Flat Plate at Angle of Attack = 20°, with Variable Reynolds Number

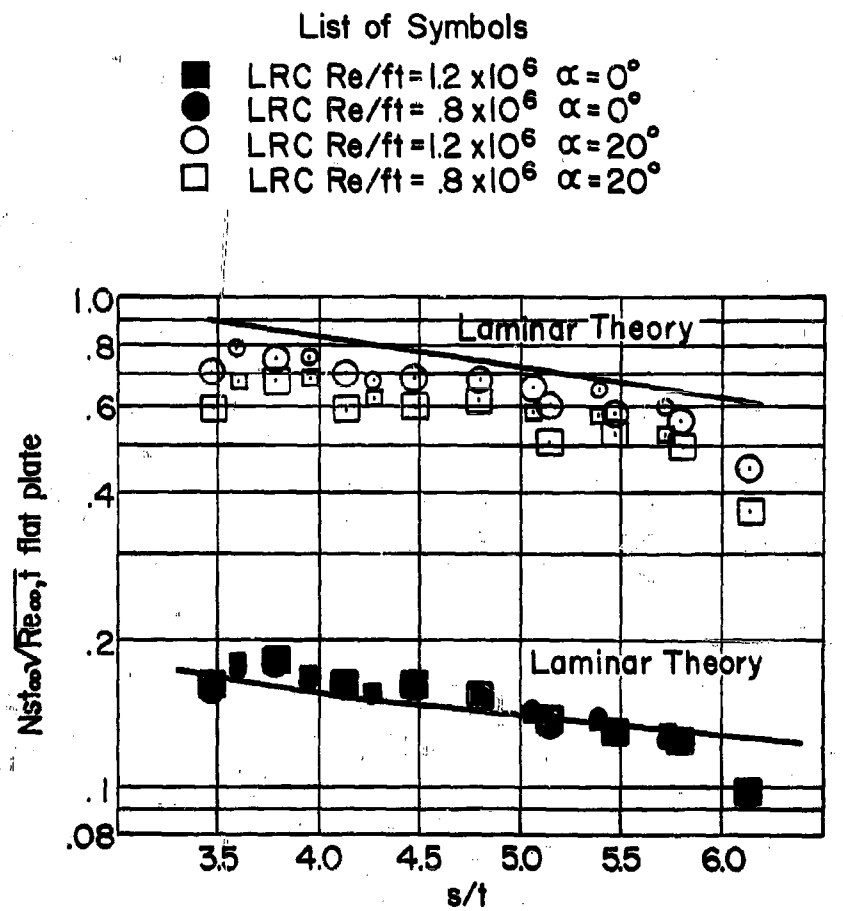


Fig. 8. Sharp Leading Edge Flat Plate at Angle of Attack = 0° and 20°, with Variable Reynolds Number

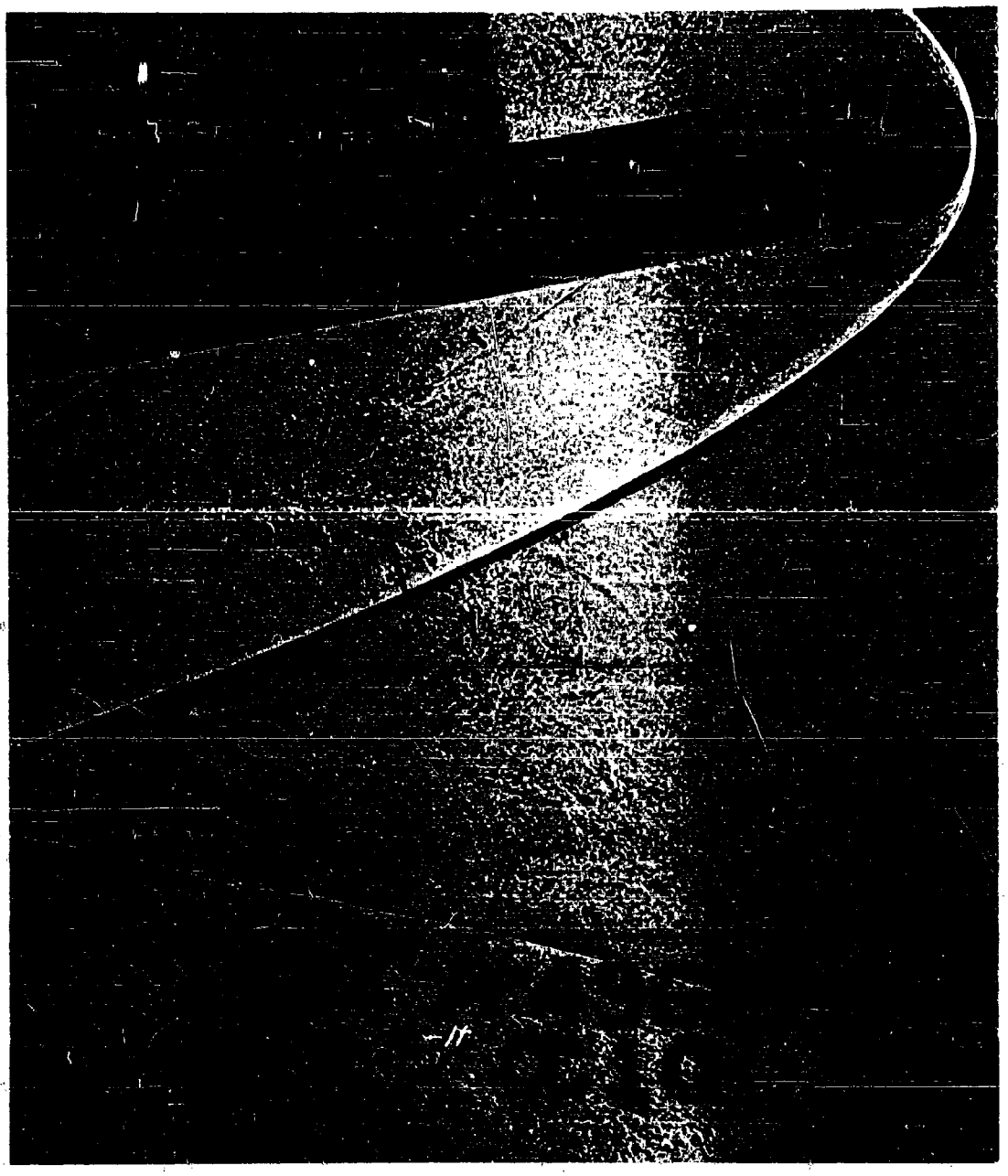


Fig. 9. Shadowgraph of Model I, Flat Plate, in Boeing Hypersonic Wind Tunnel at Angle of Attack = 11°, Reynolds Number Per Foot = 18 x 10^6, Mach 6.1

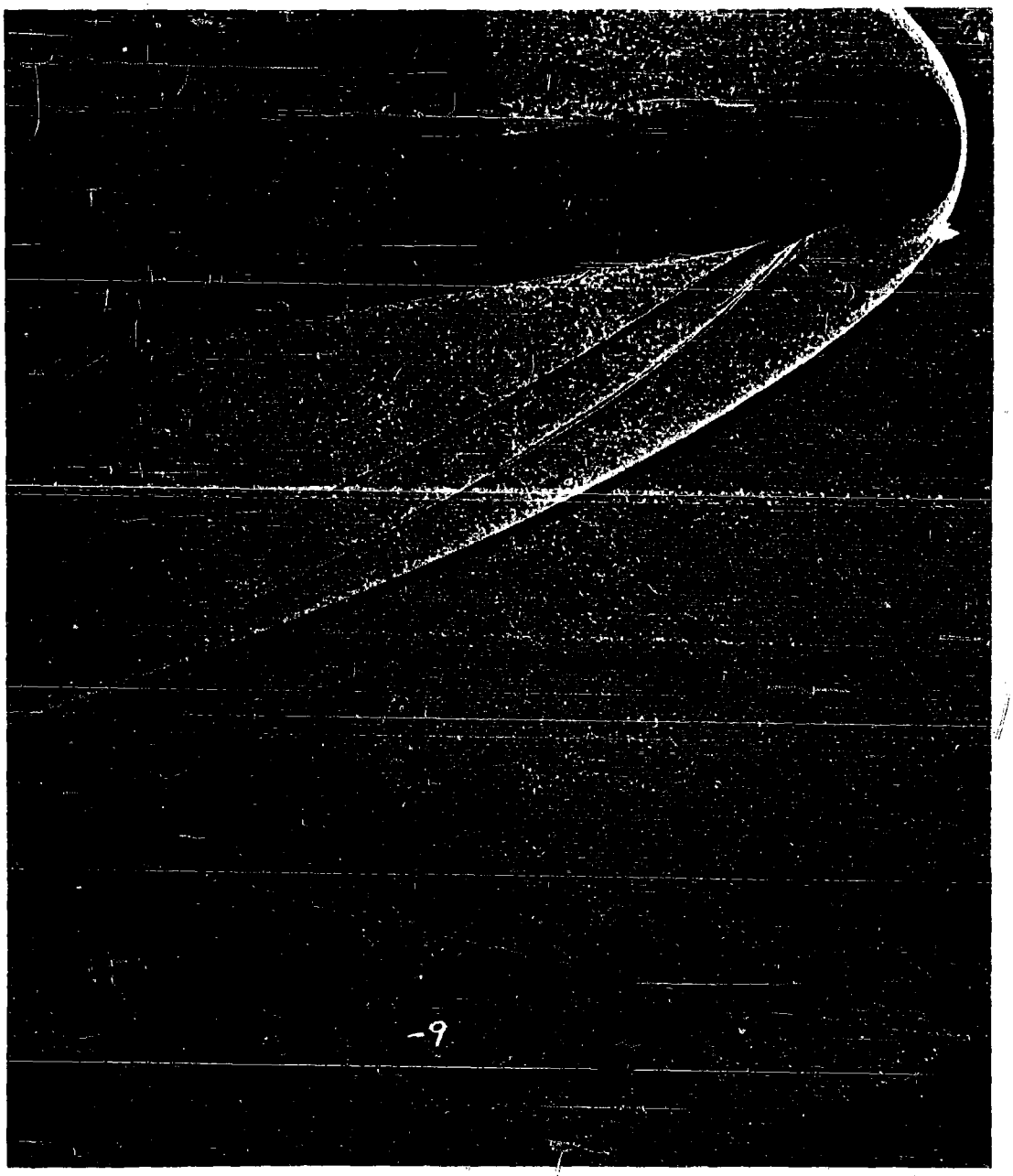


Fig. 10. Shadowgraph of Model I, Flat Plate, Tripped Flow, in Boeing Hypersonic Wind Tunnel at Angle of Attack = 11^{o}, Reynolds Number Per Foot = 18×10^{6}, Mach 6.1

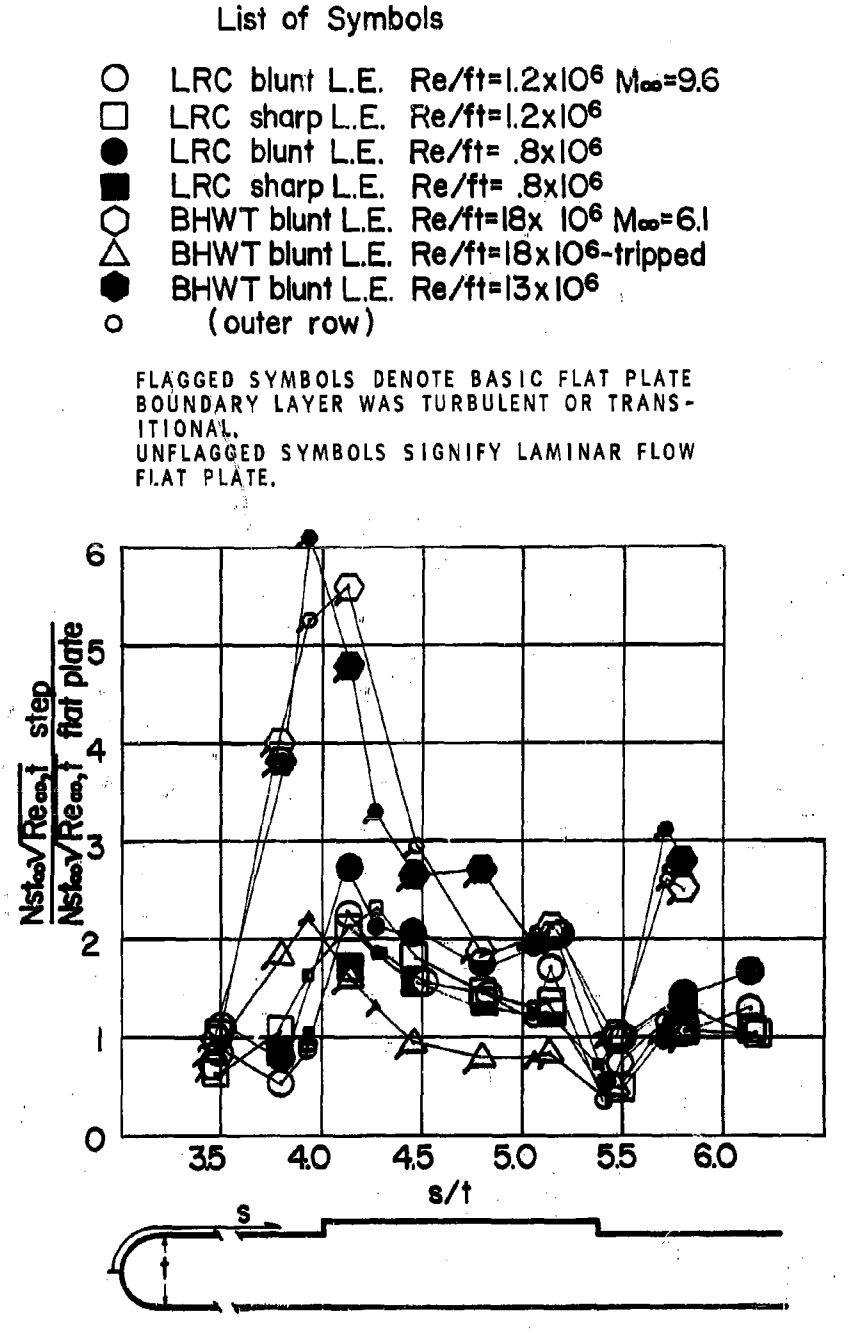


Fig. 11. Step Heating, Models I and II, at Angle of Attack = 0°, with Variable Reynolds Number

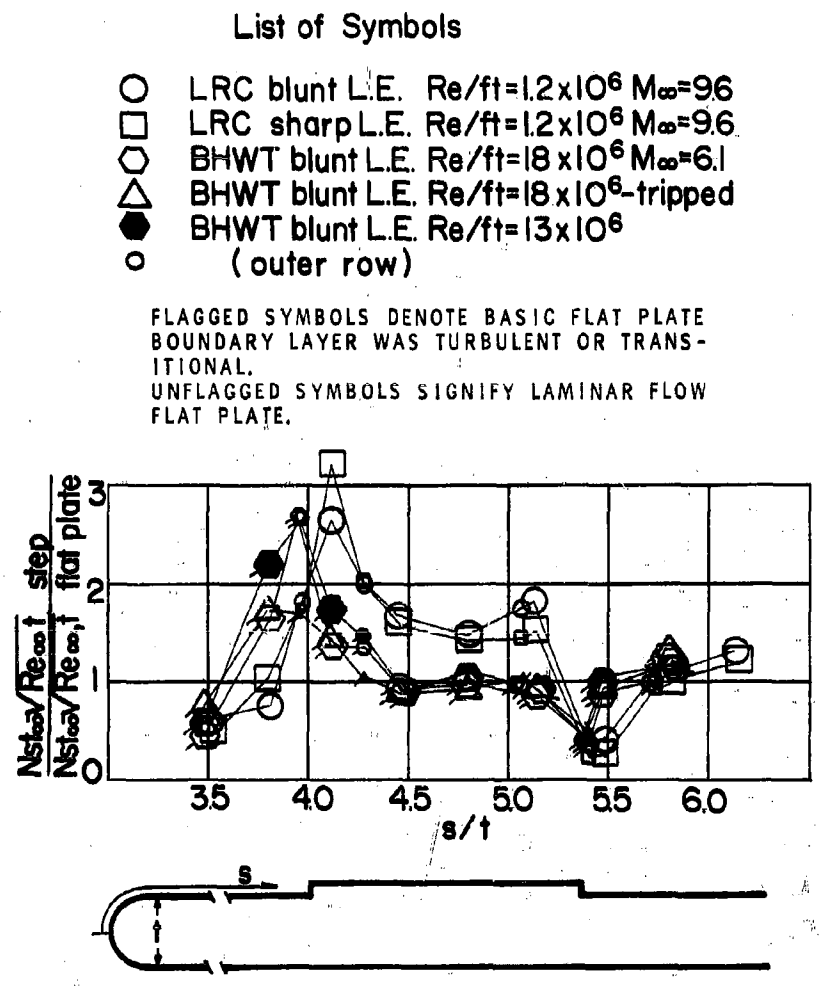


Fig. 12. Step Heating, Models I and II, at Angle of Attack = 20°, with Variable Reynolds Number

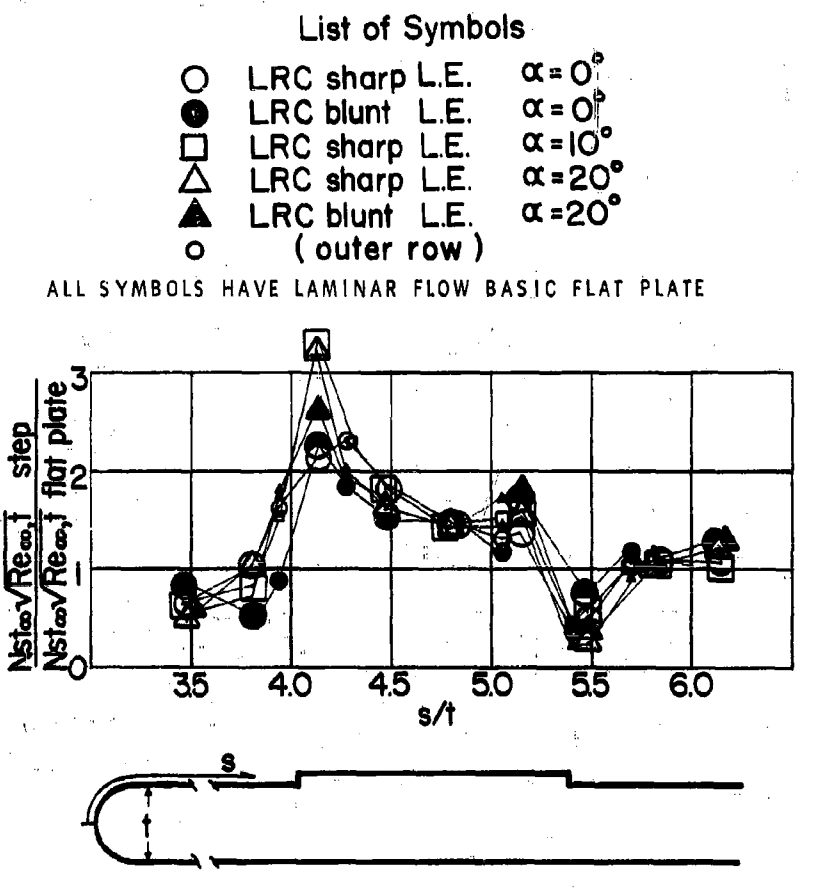


Fig. 13. Step Heating, Models I and II, at Reynolds Number Per Foot = 1.2×10^6, with Variable Angles of Attack

The tripped case, being compared with a turbulent flat plate flow, shows much lower heating ratios; they are of the same magnitude as in (2). In the LRC tests the separated flow apparently stayed laminar.

Fig. 12 shows the effect of Reynolds number at $\alpha = 20^{\circ}$. In the basic flat plate (Fig. 7) all BHWT cases have become turbulent or transitional. Thus, the BHWT runs are essentially compared with a turbulent flat plate while the LRC cases are compared with a laminar flat plate. The peak heating ratio for LRC has increased with angle of attack, conceivably due to incipient transition of the separated flow.

Fig. 13 shows the effect of angle of attack at constant Reynolds number. The corresponding flat plate plots (Fig. 8) show that laminar flow is used for comparison here. Thus, Fig. 13 appears to imply that the separated flow never became fully turbulent at the low Re of the LRC test.

The pressure distributions are plotted in Fig. 14. Peak pressure is reached ahead of the step as already found in (2).

Schlieren photographs of Model I with the step insert are shown in Figs. 15 and 16, tripped and untripped, respectively.

In summary, one can conclude that a forward facing step having a height of the same order of magnitude as the boundary layer thickness causes a localized heat flow of two to six times what would exist at the same station in the absence of the step. The highest ratio occurs when a laminar flow hits the step, separates and becomes turbulent before reattachment. The lowest ratio occurs when the separated flow remains laminar or when the flow is already turbulent before separation.

No significant conclusions can be drawn about the heat transfer downstream of the backward facing step; some of the measured values appear to be significantly higher than those of (3). This could be due to transition in the separated flow.

It is hard to ascertain the effect of boundary layer thickness from the available data because the basic flat plate flow was laminar for the low Reynolds number runs and turbulent or transitional for the high Reynolds number runs. A meaningful comparison should be based upon a purely laminar or turbulent basic flow with variable boundary layer thickness.

(3) SINE WAVES

Heat transfer ratios at $\alpha = 0$ are shown in Fig. 17 for various Reynolds numbers. Just as for the step, the sine wave caused transition in the separated flow at high Reynolds numbers. The basic flat plate runs used for comparison are of course the same as for the step, and the high heating ratios are due to the fact that a turbulent separated flow is compared with a laminar flat plate.

Fig. 18 shows the effect of variable Reynolds number at $\alpha = 20^{\circ}$. The LRC flow became transitional after separation, thus increasing the heating

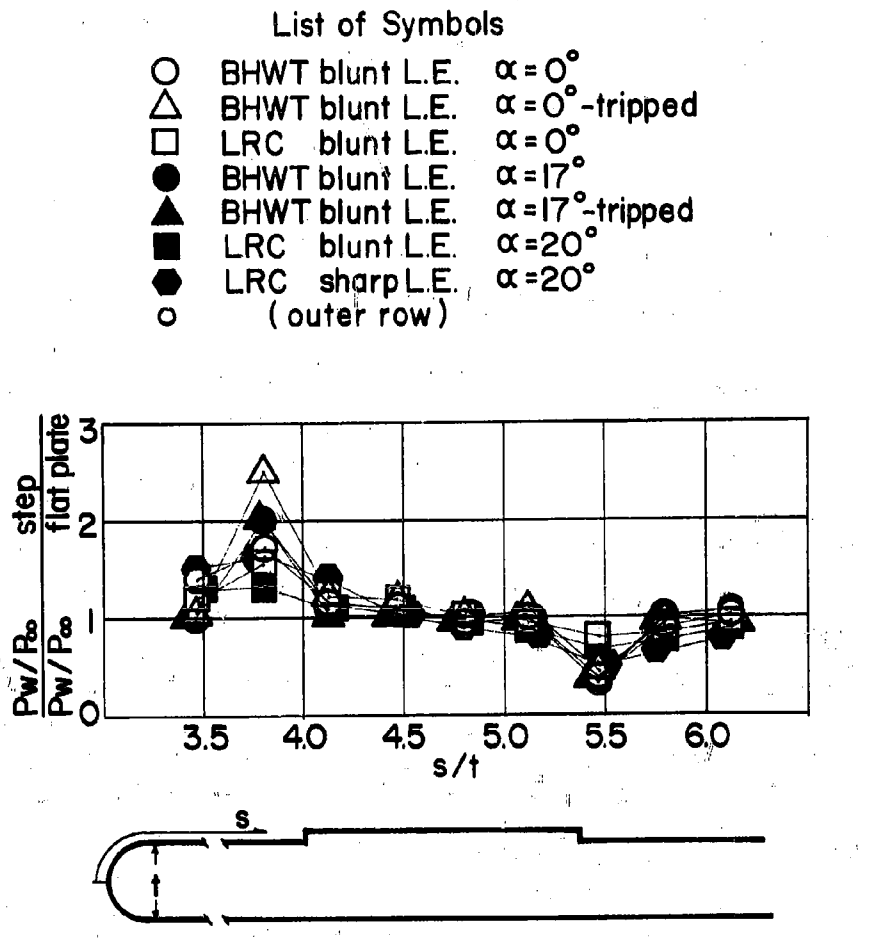


Fig. 14. Step Pressures Models I and II, at Variable Angles of Attack and Reynolds Number

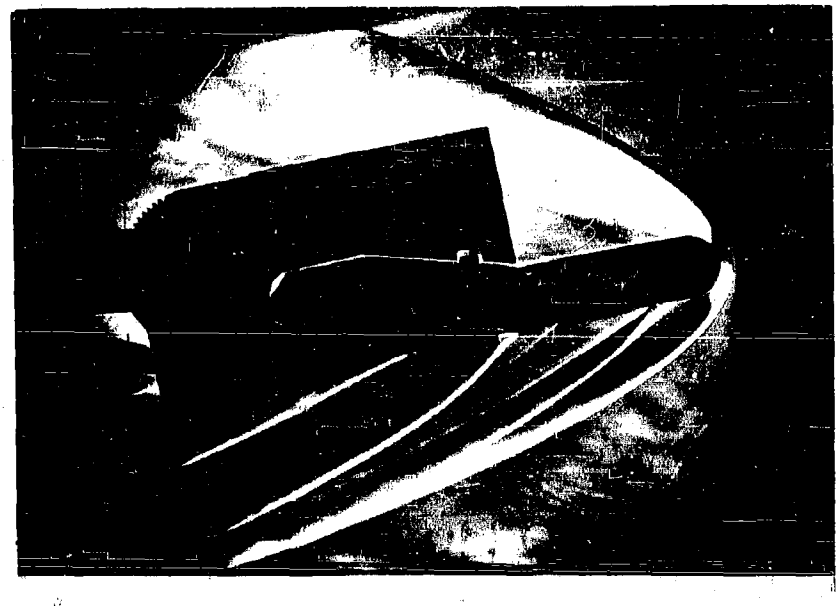


Fig. 15. Schlieren of Model I, Step with Tripped Flow, in Boeing Hypersonic Wind Tunnel at Angle of Attack = 11°, Reynolds Number Per Foot = 18×10^{6}, Mach 6.1

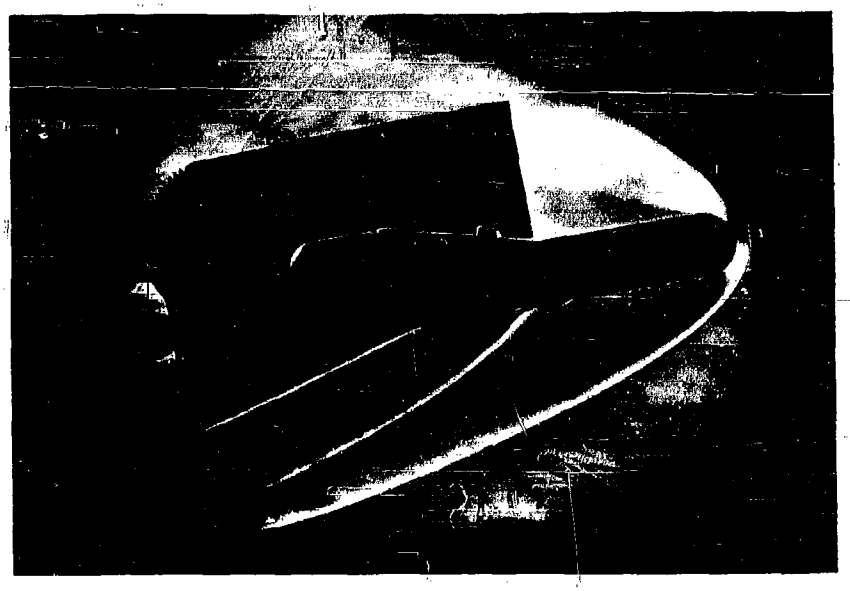


Fig. 16. Schlieren of Model I, Step, in Boeing Hypersonic Wind Tunnel at Angle of Attack = 11°, Reynolds Number Per Foot = 13×10^{6}, Mach 6.1

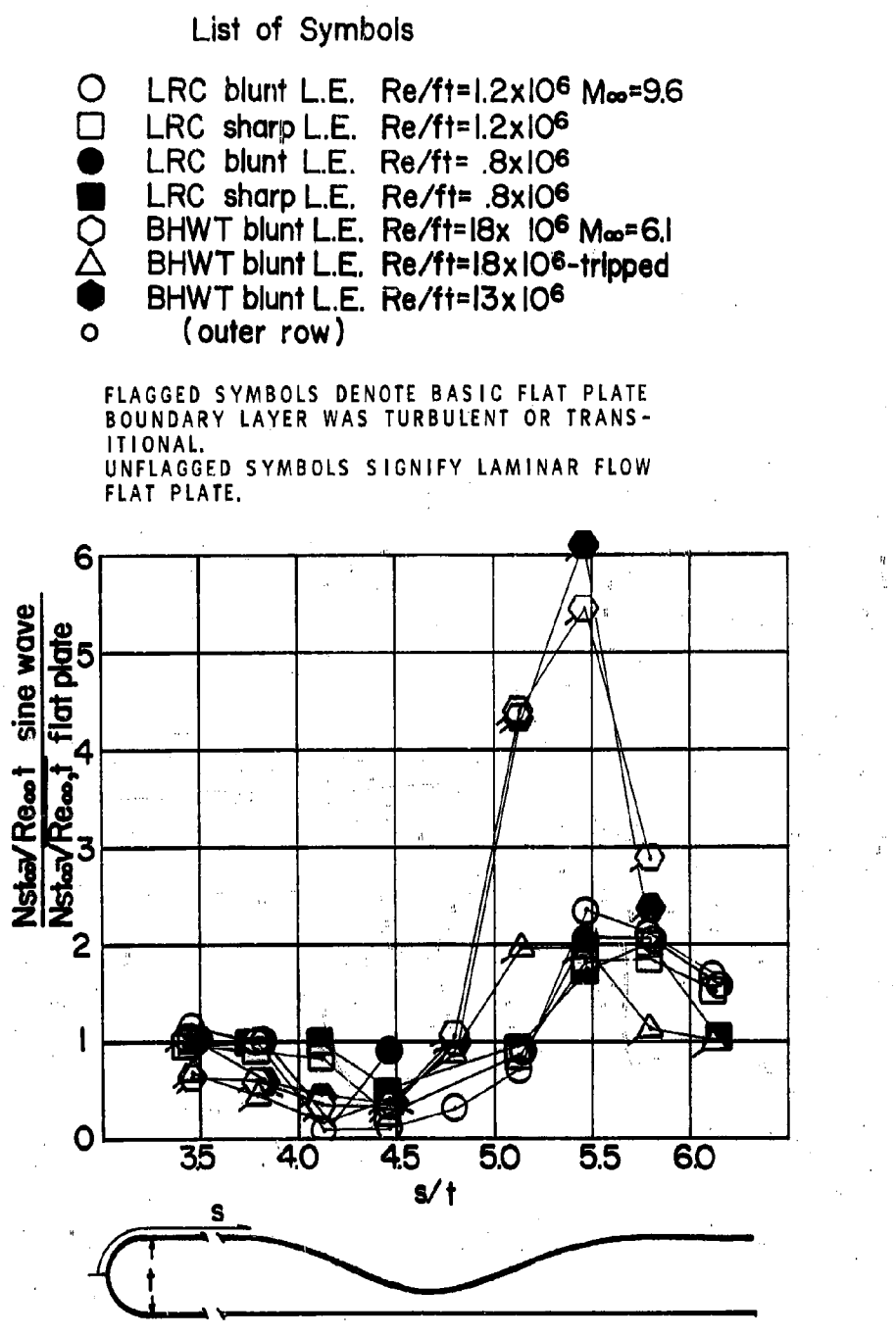


Fig. 17. Sine Wave Heating, Models I and II, at Angle of Attack = 0°, with Variable Reynolds Number

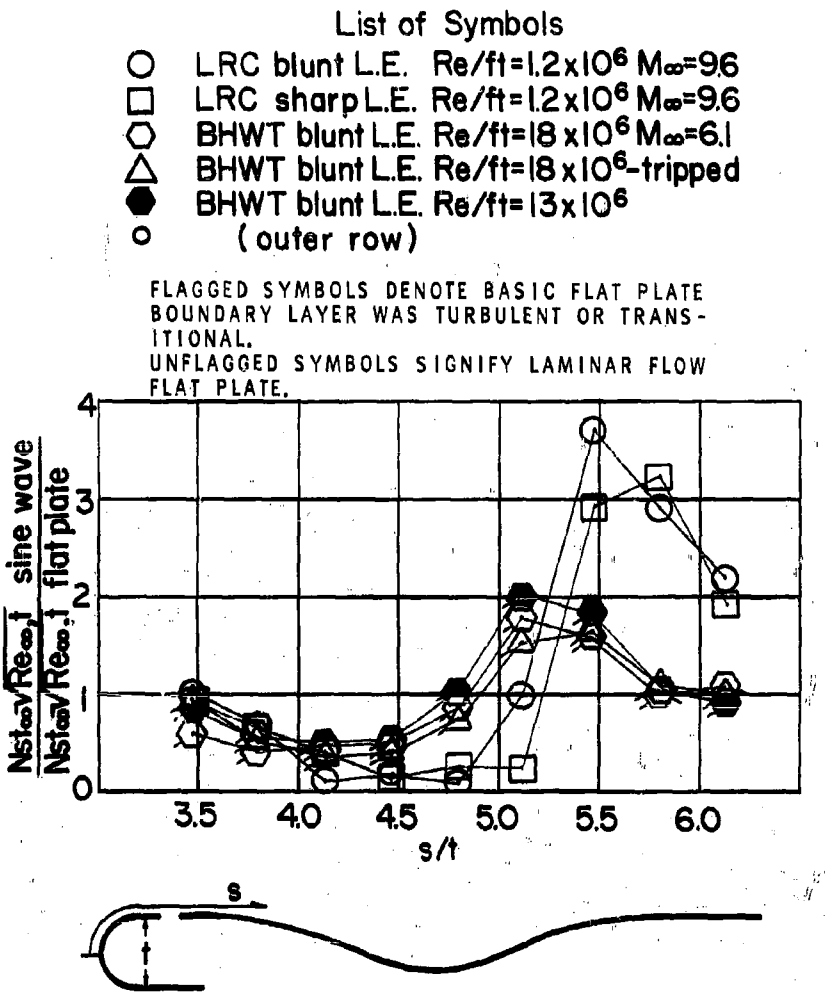


Fig. 18. Sine Wave Heating, Models I and II, at Angle of Attack = 20°, with Variable Reynolds Number

ratio since the basic flat plate remained laminar (Fig. 7). The effect of angle of attack at constant Re is shown in Fig. 19. The behaviour appears to be similar to the step case. The separated flow never became fully turbulent.

Fig. 20 shows some results obtained at AEDC for two sine waves in series, at various angles of attack. The flat plate runs were of poor quality, as mentioned previously, and the ratios were obtained on the basis of a curve faired through the flat plate data. The latter indicated laminar flow at $\alpha = 0, 10^{\circ}, 20^{\circ}$, and transitional flow at $\alpha = 30^{\circ}$.

The most significant feature is the occurrence of two successive heating peaks, the second being higher than the first. The magnitude of the second peak is open to question on account of sting effects. The absolute values of $N_{ST}\sqrt{Re}$ at the first and second peak have been plotted vs. α in Fig. 21. It is unfortunate that this test does not lend itself to a deeper study of this phenomenon; the interesting question of what the peak heating would be for a long succession of sine waves is left unanswered.

Figs. 22 and 23 show the pressure ratios. They confirm the trend shown by the heating plots, as far as location of peaks is concerned.

Figs. 24 and 25 are photographs of Model I, insert Ic, tripped and untripped, respectively.

In conclusion, it can be stated that for a sine wave corrugation with a length to depth ratio of 7.5 and a boundary layer thickness small with respect to corrugation depth, a peak heating of over twice turbulent flat plate can be expected. This agrees substantially with the conclusions of (3), Part II.

GROOVES

The small size of the grooves made it impossible to install more than one thermocouple inside each groove. Thus, the measured values may not be representative of the actual peak heating in the groove.

Figs. 26, 27, and 28 show the heat transfer ratios at various α's and Reynolds numbers. Except for the thermocouples inside the grooves, the readings at other stations were very close to flat plate readings and the scatter was considerable so that the ratios should not be regarded as too conclusive. Fig. 26 suggests that the grooves have only a slight effect on the heat transfer. This view is confirmed in Fig. 27 for angle of attack $\alpha = 20^{\circ}$. (3) gives a maximum value of 1.3 for reattachment heating ratio in a groove with $L/H = 3$. For $L/H = 1$ (our case) the ratio should be even closer to 1, in agreement with our results. Fig. 28 indicates that for low Re the heating ratio was nearly independent of angle of attack. Pressures are shown in Fig. 29. No trend can be established from these measurements. Photographs of Model Id are shown in Figs. 30, 31, tripped and untripped, respectively.

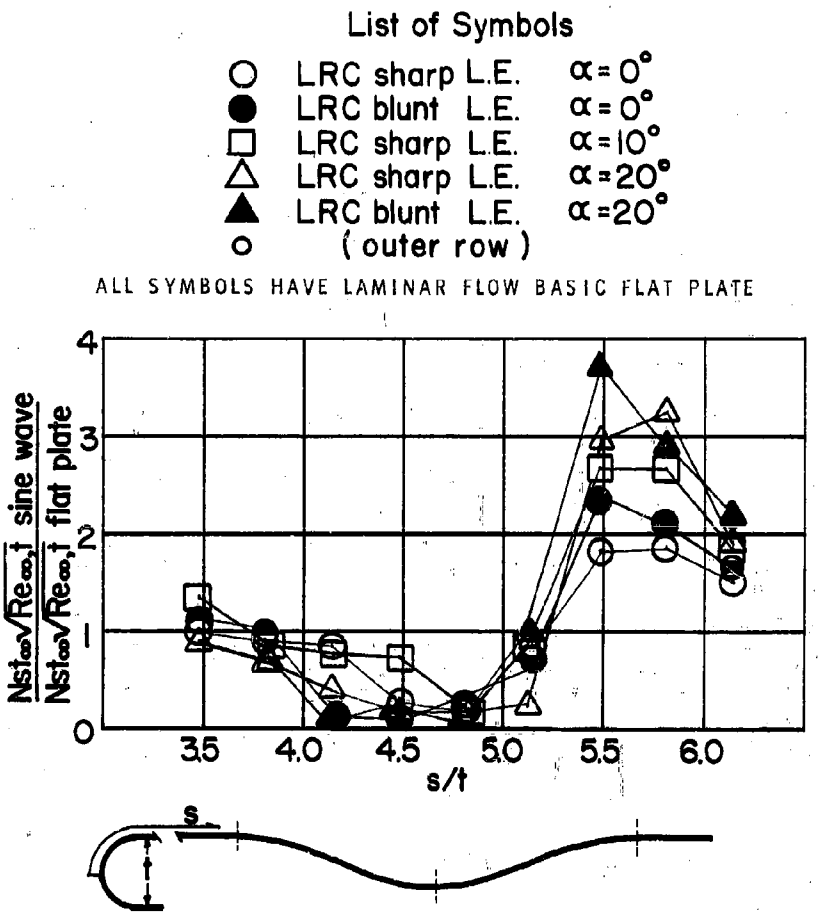


Fig. 19. Sine Wave Heating at Reynolds Number Per Foot = 1.2×10^6 and Variable Angles of Attack

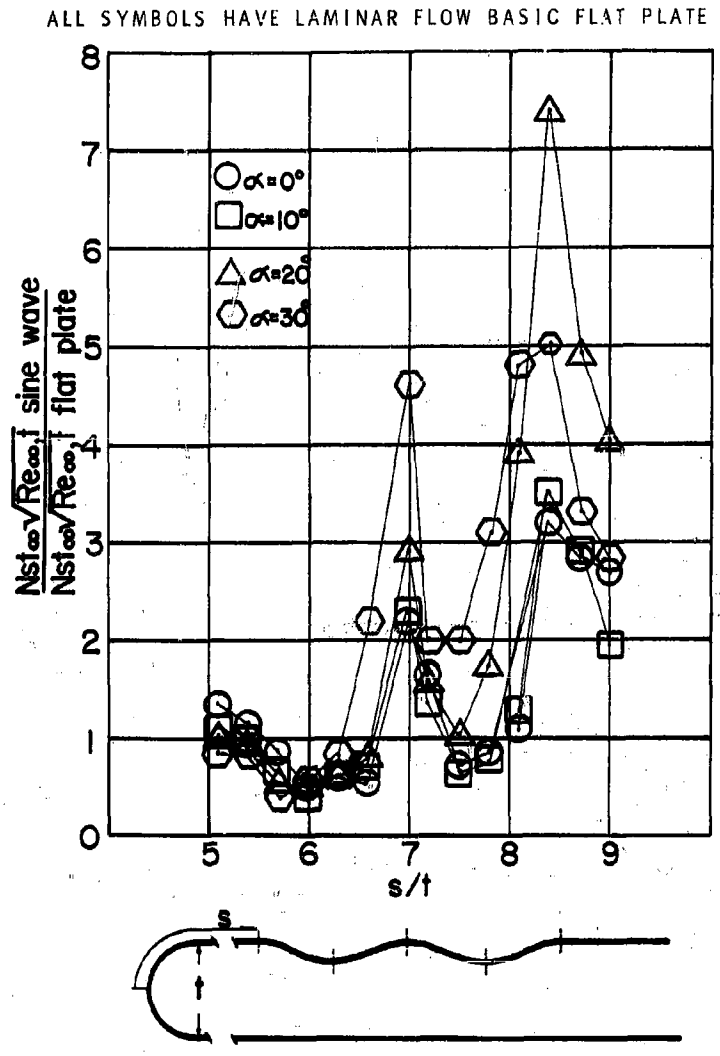


Fig. 20. Sine Wave Heating, Model III in Arnold Center "B", at Reynolds Number Per Foot = 1.5×10^6 and Variable Angles of Attack

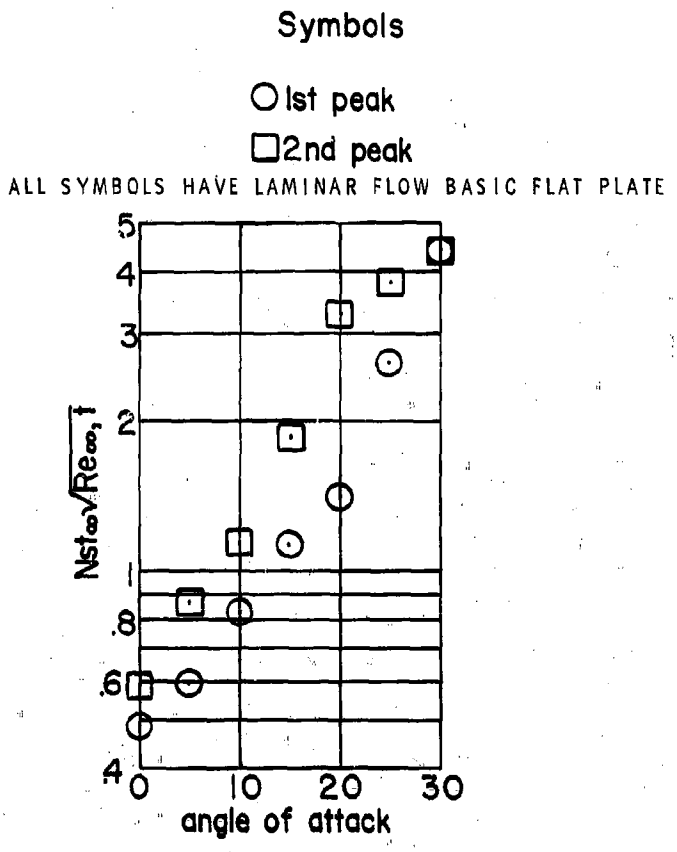


Fig. 21. Sine Wave Peak Heating, Model III in Arnold Center "B", at Reynolds Number Per Foot = 1.2×10^6 and Variable Angles of Attack

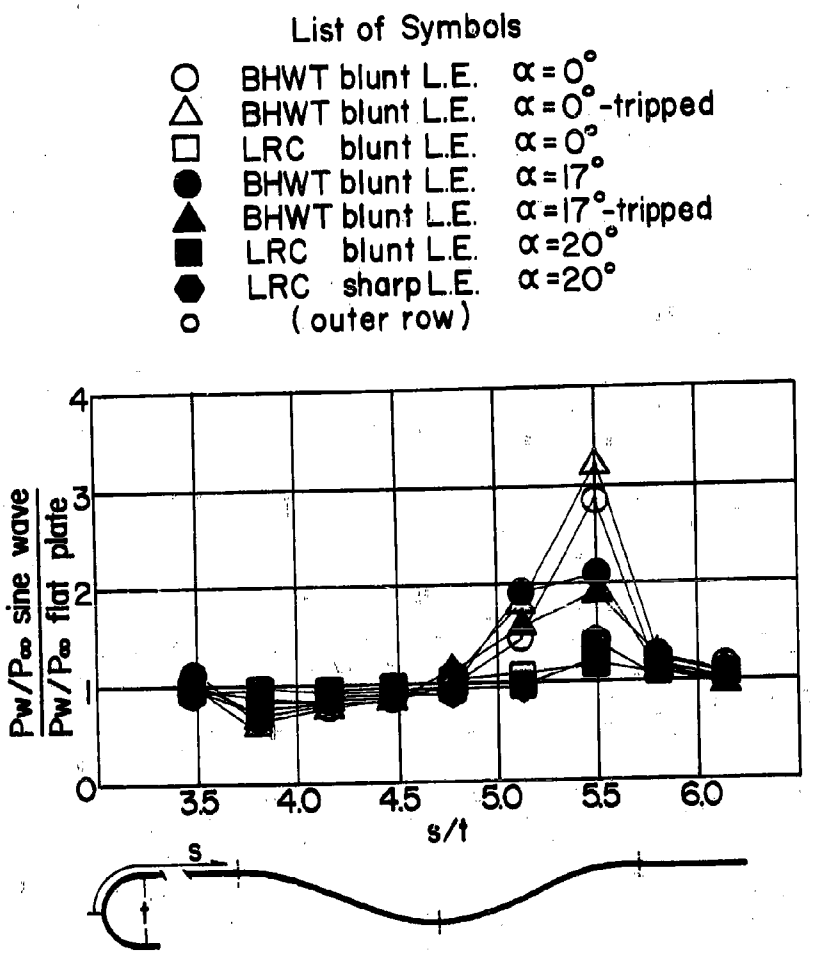


Fig. 22. Sine Wave Pressures, Models I and II at Variable Angles of Attack and Reynolds Number

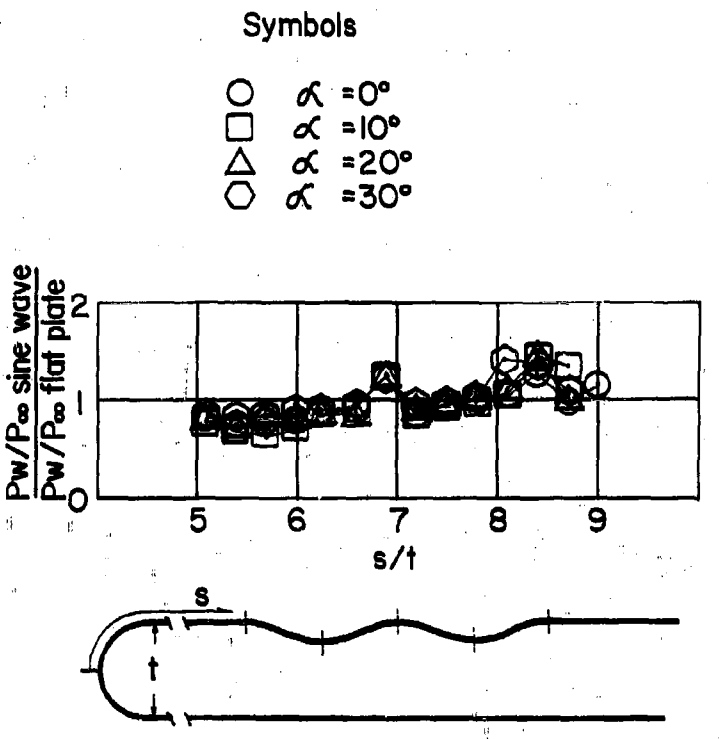


Fig. 23. Sine Wave Pressures, Model III, Mach 8 at Reynolds Number Per Foot = 1.5×10^6 and Variable Angles of Attack

Fig. 24. Schlieren of Model I, Sine Wave, with Tripped Flow, in Boeing Hypersonic Wind Tunnel at Angle of Attack = 17^{o}, Reynolds Number Per Foot = 18×10^{6}, and Mach 6.1

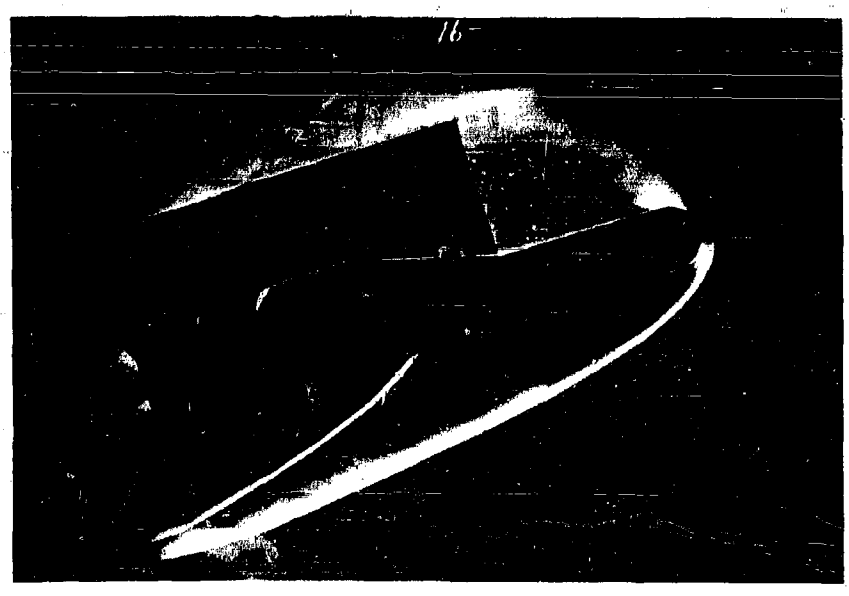


Fig. 25. Schlieren of Model I, Sine Wave, in Boeing Hypersonic Wind Tunnel at Angle of Attack = 17^{o}, Reynolds Number Per Foot = 13×10^{6}, and Mach 6.1

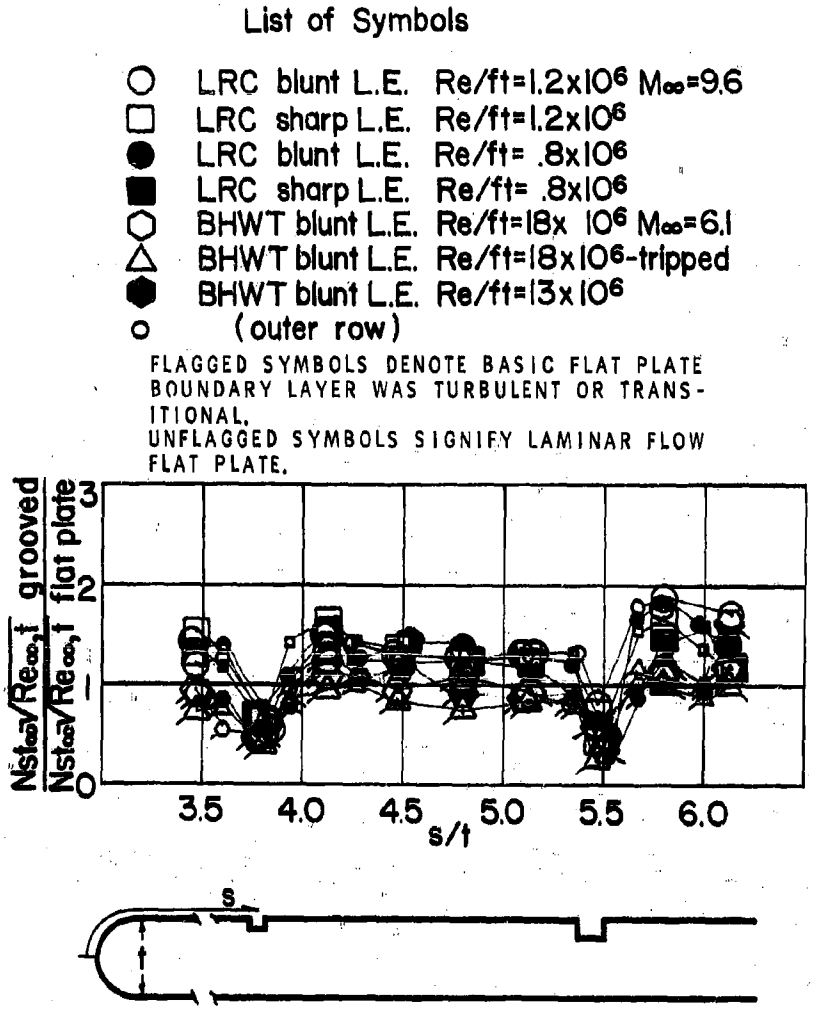


Fig. 26. Grooved Plate Heating, Models I and II, at Angle of Attack = 0°, with Variable Reynolds Number

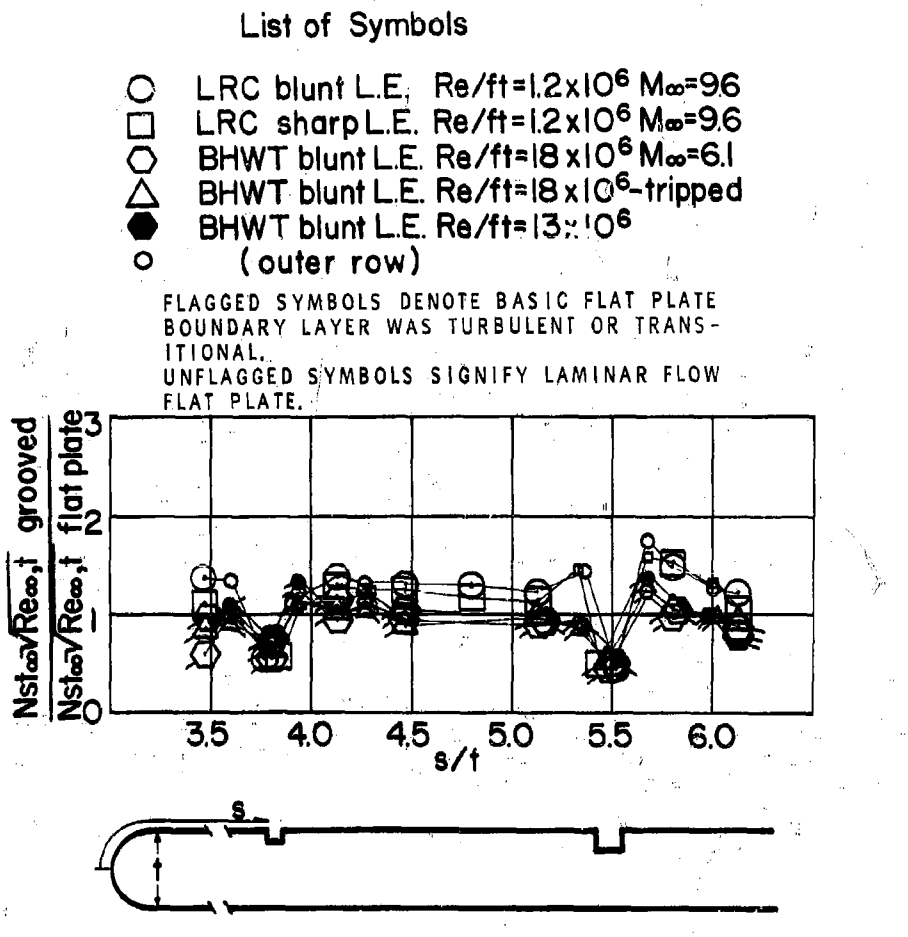


Fig. 27. Grooved Plate Heating, Models Ia-Id and II, at Angle of Attack = 20°, with Variable Reynolds Number

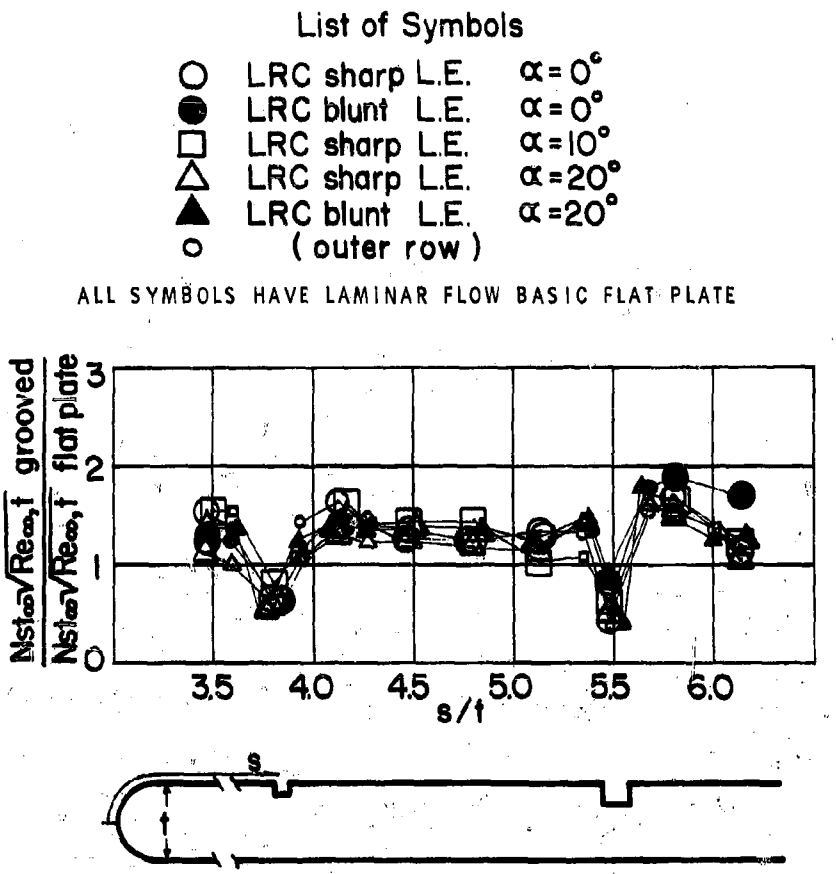


Fig. 28. Grooved Plate Heating, Models I and II, at Reynolds Number Per Foot = 1.2×10^6, with Variable Angles of Attack

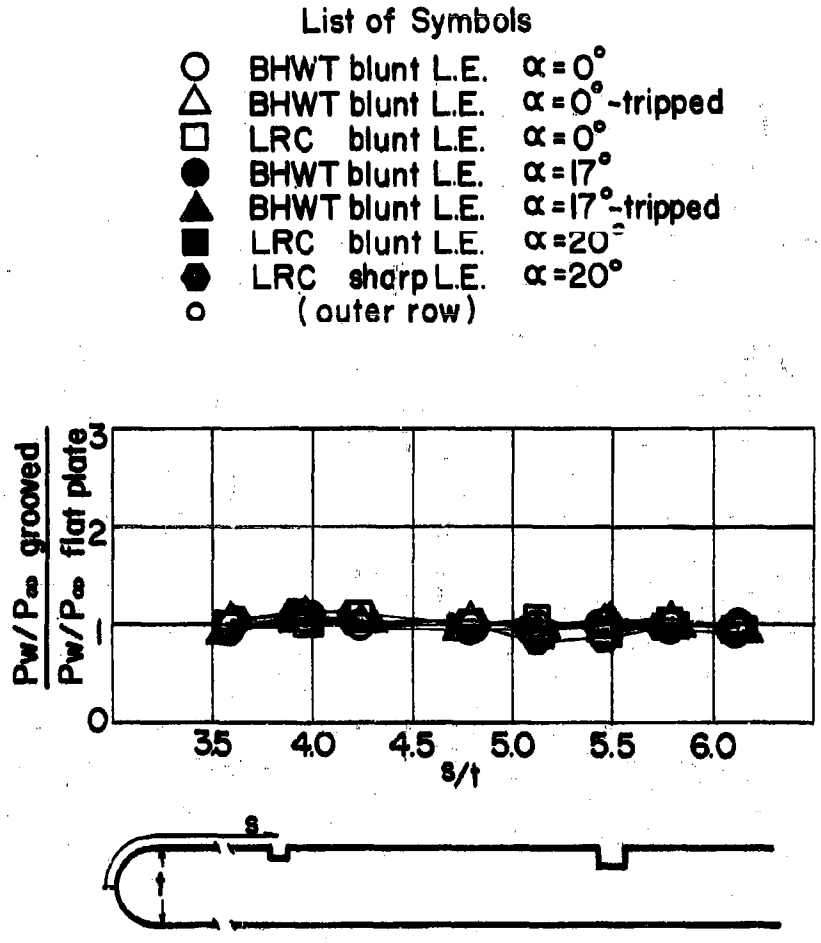


Fig. 29. Grooved Plate Pressures, Models I and II at Variable Angles of Attack and Reynolds Number

Fig. 30. Schlieren of Model I, Grooved Plate with Tripped Flow, in Boeing Hypersonic Wind Tunnel at Angle of Attack = 17°, Reynolds Number Per Foot = 18 x 10^6, Mach 6.1

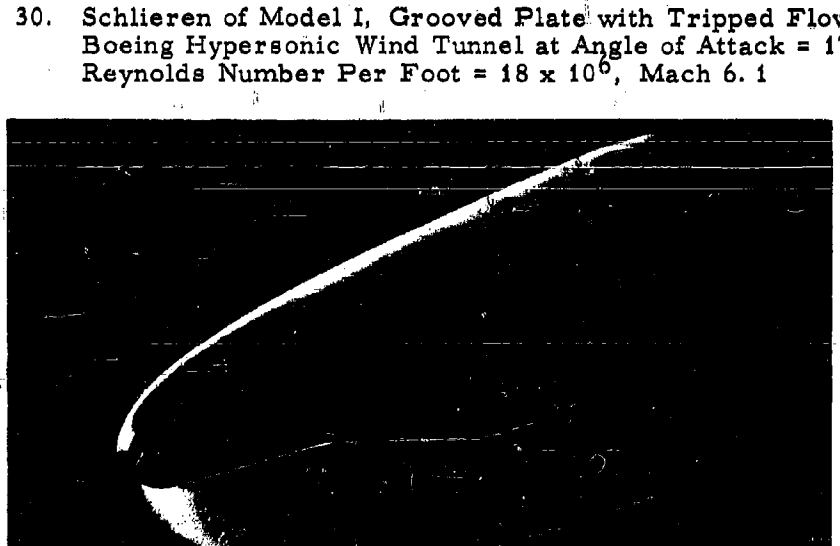


Fig. 31. Schlieren of Model I, Grooved Plate, in Boeing Hypersonic Wind Tunnel at Angle of Attack = 17°, Reynolds Number Per Foot = 13 x 10^6, Mach 6.1

CONCLUSIONS

A series of heat transfer tests on three typical surface irregularities was run. The heat transfer coefficients at every station was measured both in the presence and in the absence of irregularities, and the results were compared. Boundary layer displacement thicknesses ranged from .007 in. to .15 in.

It was found that peak heating rates occur at the reattachment points of the separated flow. Their value can be as high as six times laminar flat plate value or over twice turbulent flat plate value.

Further investigation is needed to clarify the influence of boundary layer thickness relative to irregularity depth, upon the heat transfer.

REFERENCES

1. R. S. Hiers, et al, "Heat Transfer and Pressure Distribution Tests of Several Flat Plate and Delta Wing Shapes," AEDC TN 61-40, April 1961.

2. G. E. Gadd, et al, " Heat Transfer and Skin Friction Measurements at a Mach No. 2.44 for a Turbulent Boundary Layer on a Flat Surface and in Regions of Separated Flow," NPL Report 3148, October 1958.

3. A. F. Charwat, et al, "An Investigation of Separated Flow," IAS Journal, July 1961.

4. R. Seban, et al, "Heat Transfer to Separated and Reattached Subsonic Turbulent Flows Obtained Downstream of a Surface Step," IAS Journal, December 1959.

5. H. H. Sogin, et al, "Heat Transfer in Separated Flows," ARL TR 4, January 1961.

6. J. Vasiliu, "Local and Average Heat Transfer Coefficients for a Flat Surface in Regions of Step-Induced Turbulent Separation in Supersonic Flow," Convair Report AZR-009, May 1959.

7. D. R. Chapman, "A Theoretical Analysis of Heat Transfer in Regions of Separated Flow," NASA TN 3792.

8. D. R. Chapman, et al, "Investigation of Separated Flows in Supersonic and Subsonic Streams with Emphasis on the Effect of Transition," NACA TR 1356, 1958.

9. H. H. Korst, et al, "Research on Transonic and Supersonic Flow of a Real Fluid at Abrupt Increases in Cross Section," University of Illinois ME TR 392-5, December 1959.

10. D. G. McConnel, "Free Flight Observation of Separated Turbulent Flow Including Heat Transfer up to Mach 8.5," NASA TN D-278, 1961.

11. W. B. Brower, "Leading Edge Separation of Laminar Boundary Layers in Supersonic Flow," IAS Journal, December 1961.

12. H. K. Larson, "Heat Transfer in Separated Flow," IAS Journal, November 1959.

13. A. Naysmith, "Measurements of Heat Transfer in Bubbles of Separated Flow in Supersonic Airstreams," Paper 43 of International Conference of Heat Transfer, Boulder, Colorado, August 1961.

14. R. A. Schan and J. Fox, "Heat Transfer to the Air Flow in a Surface Cavity," Paper 49 of International Conference of Heat Transfer, Boulder, Colorado, August 1961.

PROPULSION

DIELECTROPHORETIC PROPELLANT ORIENTATION IN ZERO GRAVITY

J. B. Blackmon

Douglas Aircraft Company, Inc.
Missile and Space Systems Division
Santa Monica, California

ABSTRACT

A plausible method of orienting liquid propellants in zero gravity is presented which utilizes the dielectrophoretic force induced on electrically neutral bodies in a non-uniform electric field. One of the most significant problems of propellant orientation in zero gravity is that, prior to an engine restart in space, sufficient propellant must be properly oriented for a constant flow to the engine during the restart transient. A dielectrophoretic orientation system shows promise of solving this problem. Moreover, the system has low weight and energy requirements when compared to the other methods of zero gravity propellant orientation.

We have performed experiments in which equi-density, immiscible fluids are used in experimental models to simulate conditions in a propellant tank in zero gravity. Various electrode configurations are used to collect the fluid simulating the propellant. The resulting collection times verify the theoretical calculations based on electrostatic field theory.

The possible applications of an orientation system using electric fields are discussed.

The work discussed in this paper was company sponsored.

DIELECTROPHORETIC PROPELLANT ORIENTATION IN ZERO GRAVITY

J. B. Blackmon
Douglas Aircraft Company, Inc.
Missile and Space Systems Division
Santa Monica, California

INTRODUCTION

The control of liquid propellants in zero gravity poses unique problems for future space missions which require multiple in-flight engine firings. One of the most important problems is that, prior to engine restart in zero gravity, propellant must be oriented around the tank outlet so as to provide a constant flow to the engine. If the propellant is not properly oriented around the tank outlet, slugs of gas will enter the engine causing an off-mixture ratio of the propellant and possibly an explosion.

Various orientation systems are under consideration for solving this difficulty. Some of these systems employ artificial gravity fields produced by spinning or accelerating the vehicle, and others use bellows, bladders, or piston tanks. However, these systems impose high weight penalties and energy requirements, especially when the number of engine restarts is high. Since future space missions will probably require a relatively large number of restarts, a low-energy/low-weight orientation system will be an attractive addition.

The purpose of this paper is to present the initial results of an investigation of a propellant orientation method which utilizes a relatively neglected physical effect: dielectrophoresis, the motion of electrically neutral bodies in a non-uniform electric field. Non-uniform electric fields can be produced in liquid propellant tanks by the addition of a properly designed electrode and a high-voltage power supply; the propellant tank is then a giant condenser with a high-voltage, but no current flow, between the electrodes. Propellant globules can be collected to a desired location by applying the electric field.

We have performed experiments which simulate the collection of propellant in zero gravity. A mixture of carbon tetrachloride and benzene was adjusted such that the density was equal to that of water. Since water is insoluble in carbon tetrachloride and benzene, the mixture could be used as a supporting medium for water drops. The water drops simulated propellant in zero gravity and were collected by various electrode configurations in model propellant tanks. The resulting collection times were compared with the theoretical values calculated from electrostatic field theory.

We have found that the experimental model provides a simple method of obtaining quantitative comparisons between the various electrode configurations. Moreover, the collection times with the cylindrical electrode compare favorably with theoretical values when viscous drag is considered.

Finally, the water globules simulate the physical effects of propellant collection, such as the collection trajectories, capillary forces, and random motion. Therefore, these experimental models will be useful as preliminary design tools for a dielectrophoretic propellant orientation system.

Dielectrophoresis appears promising as a method of collecting propellants because of the low weight and power requirements. The electrodes can be designed for minimum weight because they will only need sufficient structural strength to support themselves under the acceleration forces of the vehicle. Since there is no current flow, virtually the only power required is that necessary to charge the collection electrodes to a high voltage. The system is needed only intermittently prior to engine restart, and can collect common propellants over a distance of approximately one to two meters in 4 to 30 minutes at 50,000 volts. Improvements in electrode design should significantly decrease the collection times.

PROPELLANT ORIENTATION PROBLEMS IN ZERO GRAVITY

The problems of liquid propellant orientation in zero gravity are widely known throughout the aerospace industry. One of the most significant problems is that of propellant collection prior to engine restart in space. Some future space missions will require a large number of engine restarts in zero gravity. Each restart will require that sufficient propellant be available to achieve restart and provide a constant, steady flow of propellant to the engine. If the propellant is not properly oriented during restart, slugs of gas will enter the engine. The resulting propellant off-mixture ratio will cause combustion instabilities and quite possibly an explosion.

The dynamics of fluid motion under zero gravity are difficult to solve analytically and the motion of propellant in a tank in space will be essentially random because of the many disturbances involved. Each engine shutdown would induce some motion in the propellant because of the release of structural stresses caused by the vehicle acceleration. Moreover, any attitude control transients would increase the propellant motion. Recent theories on stable fluid configurations in zero gravity indicate that the propellant could be located in several different positions inside a propellant tank, depending on the previous dynamic history of shutdown, attitude control transients, etc. Therefore, without some control over propellant location, it will be virtually impossible to have propellant oriented so as to provide a constant flow to the engine during restart.

Various methods are under study for achieving reliable propellant control in space. Bellows, bladders, and piston tanks are being considered for positive control of propellant. Also, artificial gravity fields can be produced by spinning or tumbling the vehicle, or accelerating the vehicle with small restart rockets. These methods have the disadvantage of high weight and/or high energy requirements.

However, a dielectrophoretic propellant orientation system appears to present the advantages of low weight and energy, coupled with simplicity of operation. We have therefore studied the application of dielectrophoresis to propellant control in hopes that such a system will provide a solution to the problem of propellant orientation in space.

We envision light-weight electrodes located inside the main propellant tank with a high-voltage field applied. Propellant would be collected and forced to a location around the tank outlet. Both field effects and capillary attraction would be utilized to keep the collected propellant in stable equilibrium; such a system would need a minimum of energy. Typical dielectrophoretic propellant orientation systems are shown schematically in Figure 1.

DISCUSSION

Theory

An uncharged body in a non-uniform electric field experiences a force in the direction of increasing field strength. This force is the resultant of the effective attractive and repulsive forces acting on the body because of its polarization. The effect is most readily visualized if we consider the following cases. A neutral body located in a uniform electric field experiences equal forces of attraction and repulsion. These forces are caused by the separation of the centers of positive and negative charges on each molecule in the body. Effectively, then, the body acquires a negative charge on the side facing the positive electrode, and a positive charge on the side facing the negative electrode (see Figure 2).

In a non-uniform electric field, the same charge separation effect occurs in the body, but the forces of attraction and repulsion are no longer equal since the field strength varies. Then, the body experiences a force in the direction of increasing field strength. If an alternating field is used, the charge separation shifts as the field shifts and therefore the direction of the force is independent of the orientation of the field. A globule in a non-uniform electric field is depicted by Figure 3.

A clearer picture of the molecular effect of dielectrophoresis can be had by considering the differences between electrically symmetrical and unsymmetrical molecules. If the centers of electrical charge of a molecule coincide in the absence of an electric field, then the molecule is symmetrical. The dipole moment, the product of the charge times the distance of charge separation, is zero. If the molecule is unsymmetrical, the dipole moment is no longer zero and in the absence of a field, the centers of electrical charge have a permanent separation. When an electric field is applied, the charge separation or displacement polarization effect occurs in both types of molecules. In addition, the permanently polarized, unsymmetrical molecules tend to orient themselves parallel to the direction of the field. This orientation polarization or alignment, is opposed by the thermal agitation of the molecules, and a compromise is reached. In both symmetrical and unsymmetrical molecules, the charge separation or displacement polarization is due only to the field effect and is independent of temperature.

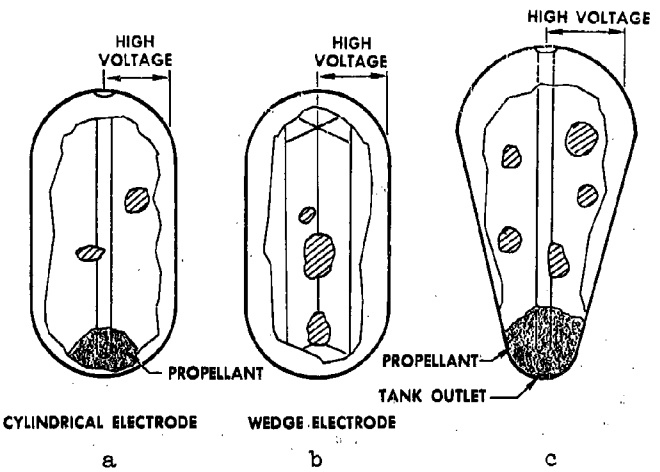


Fig. 1. Typical Orientation Systems

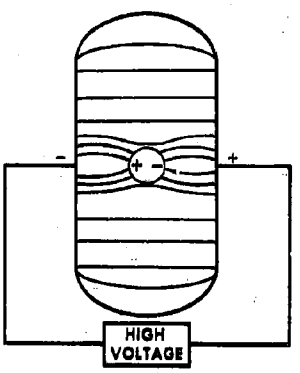


Fig. 2. Uniform Electric Field

It is interesting that electric fields at optic frequencies are varying so rapidly that the orientation polarization effect begins to decrease. This is due to the fact that the molecular motion is slow relative to the variation in the field at optic frequencies. Moreover, the displacement polarization effect begins to decrease when the field frequency is in the region of the characteristic vibrating frequency of the electrons in the molecules. However, neither of these effects occurs at normal alternating current frequencies and therefore the polarization, and hence dielectric constant, will not vary with the field frequency.

When we consider the body in an electric field on a macroscopic scale, the dielectric constants of the supporting medium and the body become prime parameters. If the dielectric constant of the body is greater than that of the medium, then the body would be collected by a non-uniform electric field as shown by Figure 3. If the dielectric constant of the body is less than the surrounding medium, then the body will be displaced by the medium, and will migrate to the outside. In the case of liquid propellant in zero gravity, the supporting medium is a gas, with a dielectric constant of approximately 1.0. The dielectric constants of the propellant can vary from 1.25 for liquid hydrogen to 53 for hydrazine. Therefore, liquid propellant globules in space vehicles will always tend to move in the direction of increasing field strength.

Propellant Collection Equations

We can gain a better understanding of this phenomena by deriving the equations of motion for a small globule in a non-uniform field. Consider the liquid globule of dielectric constant K" suspended in the base liquid of dielectric constant K', in the non-uniform electric field formed by two concentric cylinders of inner radius a and outer radius b (Figure 3). If the globule is small compared to the inner radius a, we can assume as a first-order approximation that the globule does not distort the field and only the surface stresses resulting from the normal electric displacement vector, D_N, will be considered.

Since there is no free charge on the interface, Gauss' electric flux theorem for a non-homogeneous medium states that

$$D'dS = D''dS \text{ or } D' = D'' \tag{1}$$

The stress, or force per unit area, is given in terms of the electric displacement, D, as

$$F = \frac{D^2}{2\epsilon} \tag{2}$$

It should be noted here that since the globule is in a non-uniform field, the electric field intensity at the second interface is less than that at the first interface. Also, we are assuming an average charge distribution located at the positions r_1 and r_2. The radii r_1 and r_2 are related to the bubble position and size by the following argument.

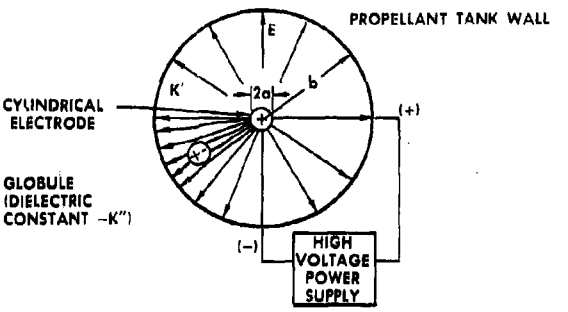


Fig. 3. Top View of Cylindrical Electrode

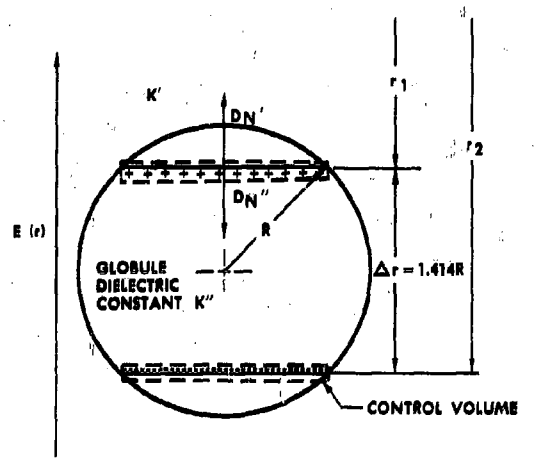


Fig. 4. Spherical Globule with Effective Charge Distribution and Geometry

We can make the engineering approximation that the globule charge distribution is nearly the same as that of a sphere in a uniform field. Therefore, the charge distribution is a cosine function, and the average effective charge is distributed on the base of the spherical segment of the circle located at .707 of the radius from the center (see figure 4). We can now assume that the effective charge is distributed across the flat surface at .707 R.

The stress on a surface due to the normal components of the electric displacement must be the difference in the stresses on both sides of the small control volume drawn around the interface (see figure 4). Therefore, the net stress on interface number 1 is given by:

$$F(r_1) = -\left(\frac{D_1'^2}{2\epsilon'} - \frac{D_1''^2}{2\epsilon''}\right) \tag{3}$$

Applying Gauss' theorem and the relationship $K = \epsilon/\epsilon_v$, we have:

$$F(r_1) = -\frac{D_1'^2}{2\epsilon_v}\left(\frac{K'' - K'}{K'K''}\right) \tag{4}$$

The expression for the stress at the second interface is the same as above except for a sign change because of repulsion from the central charge.

$$F(r_2) = \frac{D_2'^2}{2\epsilon_v}\left(\frac{K'' - K'}{K'K''}\right) \tag{5}$$

The total accelerating force per unit area is the algebraic sum of the interface stresses.

$$\Delta F(r) = -\frac{K'' - K'}{2\epsilon_v K' K''}\left(D_1'^2 - D_2'^2\right) \tag{6}$$

The electric displacement is related to the field intensity by:

$$D = \epsilon E \tag{7}$$

or:

$$D_1' = \epsilon' E(r_1)$$

$$D_2' = \epsilon' E(r_2)$$

The length of the globule, Δr, is small compared to r. Therefore, we can expand the expression for $E(r_2)$ in a Taylor series expansion, neglecting higher order terms.

$$E(r_2) = E(r_1) + \frac{\partial}{\partial r} E(r)\,\Delta r \tag{8}$$

Substituting equations (7) and (8) into equation (6), we obtain

$$\Delta F(r) = \frac{K'' - K'}{K''} K' \epsilon_v \quad E(r) \frac{\partial E(r)}{\partial r} \Delta r \tag{9}$$

For a cylindrical condenser with an isotropic homogeneous medium, the voltage can be expressed as

$$V(r) = V_a - V_a \frac{\ln (\frac{r}{a})}{\ln (\frac{b}{a})} \tag{10}$$

where r = radial distance
a = radius of the axial, high-voltage rod
b = inner radius of the tank.

This expression is assumed for the voltage distribution in the cylinder for the case of a globule with a different dielectric constant. There is little error introduced by this assumption if the globule size is small.

We know that $E = -$ grad V, or

$$E(r) = \frac{V_a}{r \ln \frac{b}{a}} \tag{11}$$

Also,

$$\frac{\partial E(r)}{\partial r} = - \frac{V_a}{\ln (\frac{b}{a}) r^2} \tag{12}$$

Substituting equations (11) and (12) into (9),

$$\Delta F(r) = \frac{(K'' - K') K'}{K''} \epsilon_v \frac{V_a^2 \Delta r}{r^3 (\ln \frac{b}{a})^2} \tag{13}$$

Considering the motion of the particle in the field and ignoring viscous drag forces for a first approximation, we have

$$\Delta F(r) \cdot A = M \frac{dv}{dt} \tag{14}$$

where $M = \rho V$ and A is the area of effective charge distribution.

From the geometry of the sphere and the definitions of A and Δr, we know that

$$\Delta r = 2 \cdot (.707) R \tag{15a}$$

and

$$A = \frac{\pi R^2}{2} \tag{15b}$$

$$\therefore V = \frac{4\ A\ \Delta r}{3\ (.707)} \tag{16}$$

Substituting the above into equation(14) we have

$$\frac{dv}{dt} = \frac{3\ (.707)\ \Delta F(r)}{4\ \Delta r\ \rho} \tag{17}$$

From the chain rule, we know that

$$\frac{dv}{dt} = \frac{dv}{dr} \cdot \frac{dr}{dt} = v\ \frac{dv}{dr} \tag{18}$$

Therefore,

$$\int_0^V v dv = -\frac{3(.707)(K''-K')\ V_a^{\,2}\ \int_{r_o}^{r} \frac{dr}{r^3}}{4\ K''\ \rho\ (\ln \frac{b}{a})^2} \tag{19}$$

Integrating,

$$V(r) = \sqrt{\frac{3(.707)(K''-K')K'\epsilon_v\ V_a^{\,2}(r_o^{\,2}-r^2)}{4\ \rho(\ln \frac{b}{a})^2\ r_o^{\,2}\ r^2}} \tag{20}$$

which yields

$$V(r) = \frac{V_a}{2\ r_o r\ \ln \frac{b}{a}} \sqrt{\frac{3(.707)(K''-K')K'\epsilon_v\ (r_o^{\,2}-r^2}{K''\ \rho}} \tag{21}$$

To obtain the collection time, we integrate once more

$$\int dt = \int_{r_o}^{r} \frac{dr}{V(r)}$$

$$t = \frac{2\ r_o\ \ln \frac{b}{a}}{V_a \sqrt{\frac{3(.707)(K''-K')K'\epsilon_v}{K''\ \rho}}} \int_{r_o}^{r} \frac{r\ dr}{r_o^{\,2}-r^2} \tag{22}$$

and

$$t = \frac{2\, r_o \ln \frac{b}{a} \sqrt{r_o^2 - r^2}}{V_a \sqrt{\frac{3(.707)(K'' - K')\, K' \epsilon_v}{K'' \rho}}} \tag{23}$$

Equation (23) is the theoretical collection time of a small globule in a non-uniform electric field formed by two concentric cylinders of radii a and b; the initial globule position is given by r_o. This equation will be used to calculate collection times in space vehicles; viscous drag is neglected.

It has been shown in our model experiments with the viscous fluids that the magnitude of the globule acceleration, or its equivalent inertial force, is much smaller than the field force and drag. Therefore, the viscous drag and field force can be assumed equal in our experiments. This approximation is necessary in order to obtain a simple solution to the equation. Equating the viscous drag expression from Stokes' law to the field forces, we have

$$6 \pi \eta v R + \Delta F \cdot A = 0 \tag{24}$$

Substituting for $\Delta F(r)$ and solving

$$V = \frac{(K'' - K')\, K' \epsilon_v\, V_a^2\, \Delta r\, A}{K'' \left(\ln \frac{b}{a}\right)^2 R\, r^3\, 6 \pi \eta} \tag{25}$$

we again use the relationships for Δr, A, and R.

Therefore,

$$V = \frac{(K'' - K')\, K' \epsilon_v\, V_a^2 (.707)\, R^2}{6\, K''\, \left(\ln \frac{b}{a}\right)^2 \eta\, r^3} \tag{26}$$

Integrating with respect to time, we have

$$t = \frac{3\, K''\, \left(\ln \frac{b}{a}\right)^2 \eta\, (r_o^4 - r^4)}{2(.707)(K'' - K')\, K' \epsilon_v\, V_a^2\, R^2} \tag{27}$$

Equation (27) gives the collection time for droplets in a viscous medium with a non-uniform electric field formed by two concentric cylinders of radii a and b. Both equations (23) and (27) hold only for the concentric cylindrical electrode and tank configurations. However, it will be shown that when coupled with experimental data, these equations will enable us to calculate the collection time in space for various other, more complicated, electrode designs.

EXPERIMENT

Purpose

Our experimental program was designed to serve three purposes. First, we needed comparative data on the various electrode shapes. By comparing the collection times of the various electrodes under the same conditions, we could determine which electrode configurations are most effective in propellant collection. This is necessary because some electrode configurations produce electric fields which are difficult to solve analytically. Therefore, we compare the collection times of the complicated electrodes to the collection time with the cylindrical electrode. Then we calculate the collection time for the real case in space with the cylindrical electrode and apply the electrode effectiveness, or the ratio of collection times, to determine the collection time in space for the more complex electrode shapes. The results obtained are approximate but are sufficient for preliminary comparisons. Of course, it is important in these comparisons that all parameters such as voltage, droplet size, and tank size remain constant.

The second purpose of the experiments was to simulate the expected dynamic conditions of propellant in zero gravity. We felt that this goal was realistic since the motion of large water globules in our supporting fluid closely resembled the motion of wetting liquids in zero gravity tests made on KC-135 aircraft flights. Moreover, we felt that the particular capillary attraction effects of the water on the various electrode shapes would be typical of the real case in space. Unfortunately, even the de-ionized, distilled water used in our experiments was too highly conductive for some of our experiments, and on occasion, electrodes would short out. We did manage to simulate a few cases of propellant collection with large amounts of water. These results are described under the "Propellant Collection" discussion.

The third purpose of the experiments was to reveal problem areas we did not initially envision. We were constantly faced with a number of problems which made the collection of reliable data a difficult, time-consuming task. First, it was absolutely necessary to keep the fluids as pure as possible throughout the experiments. Impurities drastically reduced the specific resistivity of our supporting fluid, and therefore created the detrimental effect of electric and ionic current flow to the electrodes. We feel that this is an experimental problem and should not normally occur in space vehicle applications.

Another problem we faced was that a net charge could be induced on the water drops in the vicinity of the electrodes. This effect occurred at voltages above 2000 to 3000 volts with our apparatus, and became more significant as the voltage increased. Each time the drop would gain a net charge, the electrophoretic effect (motion of a charged body in an electric field) would repel it from the central electrode. As the drop moved out, it lost its charge, and the dielectrophoretic effect would again pull the drop radially inward.

The ionization effect in a region of high electric field intensity appears to be one of our most significant experimental problems. This effect is described more fully under "Electrophoretic Phenomena."

Experimental Setup

Before proceeding to a discussion of our experimental results, a brief description of the experimental setup is necessary. Our experiments were made in a covered laboratory jar 10 inches in diameter. A cylindrical aluminum wall was fitted into the jar to simulate a propellant tank wall. Electrodes were supported from the jar covering. The jar was partially filled with a mixture of spectrograde carbon tetrachloride and benzene, adjusted such that the density was equal to that of water. A pipette was used to place a 0.1 ml or 0.2 ml water drop in the mixture (see Figure 5a). Spectrograde carbon tetrachloride and benzene were chosen for the experiment because their relatively high specific resistivity (on the order of 10^{+16} ohm^{+1} cm^{+1}) made them excellent insulators. Until the fluids became contaminated with impurities, only a negligible current flowed between the electrodes. However, impurities lowered the specific resistivity of the fluids and caused an ionic current flow. The current flow resulted in electrophoresis effects.

Two power supplies were used. We first used an AC power supply consisting of an ignition transformer and Variac; the output voltage was calculated from the transformer turns ratio and the primary voltage and measured with a voltmeter at the lower voltages. The other power supply was a high-voltage direct current insulation tester with a calibrated voltage and current readout. Both power supplies had a voltage output of approximately 15,000 volts, but the maximum voltages were not used in our experiments.

There was the danger of an explosion because of the high voltages used and the benzene present in the tank model. Moreover, both benzene and carbon tetrachloride are poisonous. Therefore, the equipment and observers were separated from the tank model by a wall and observations were taken through a safety glass window. A mirror was placed above the model as shown in Figure 5b so that observations and photographs could be easily and safely made.

After allowing all fluid motion to come to rest, the droplet location in the tank was determined. Then, the voltage was applied, and an electric timer started. The timer was stopped as the droplet made contact with the inner electrode. The photographs of Figures 6 to 8 illustrate a typical collection.

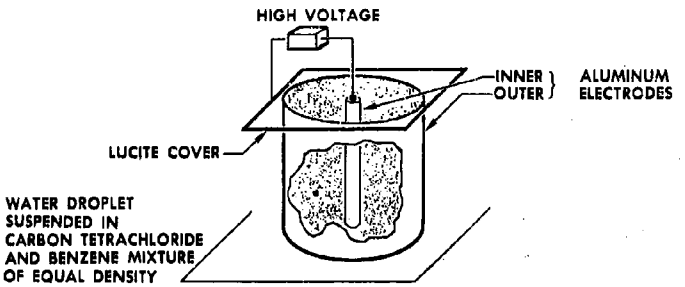


Fig. 5a. Experimental Model Cylindrical Electrode

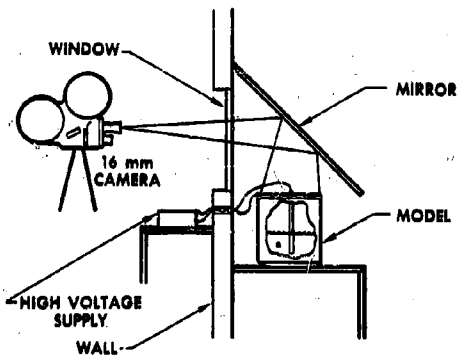


Fig. 5b. Experimental Setup

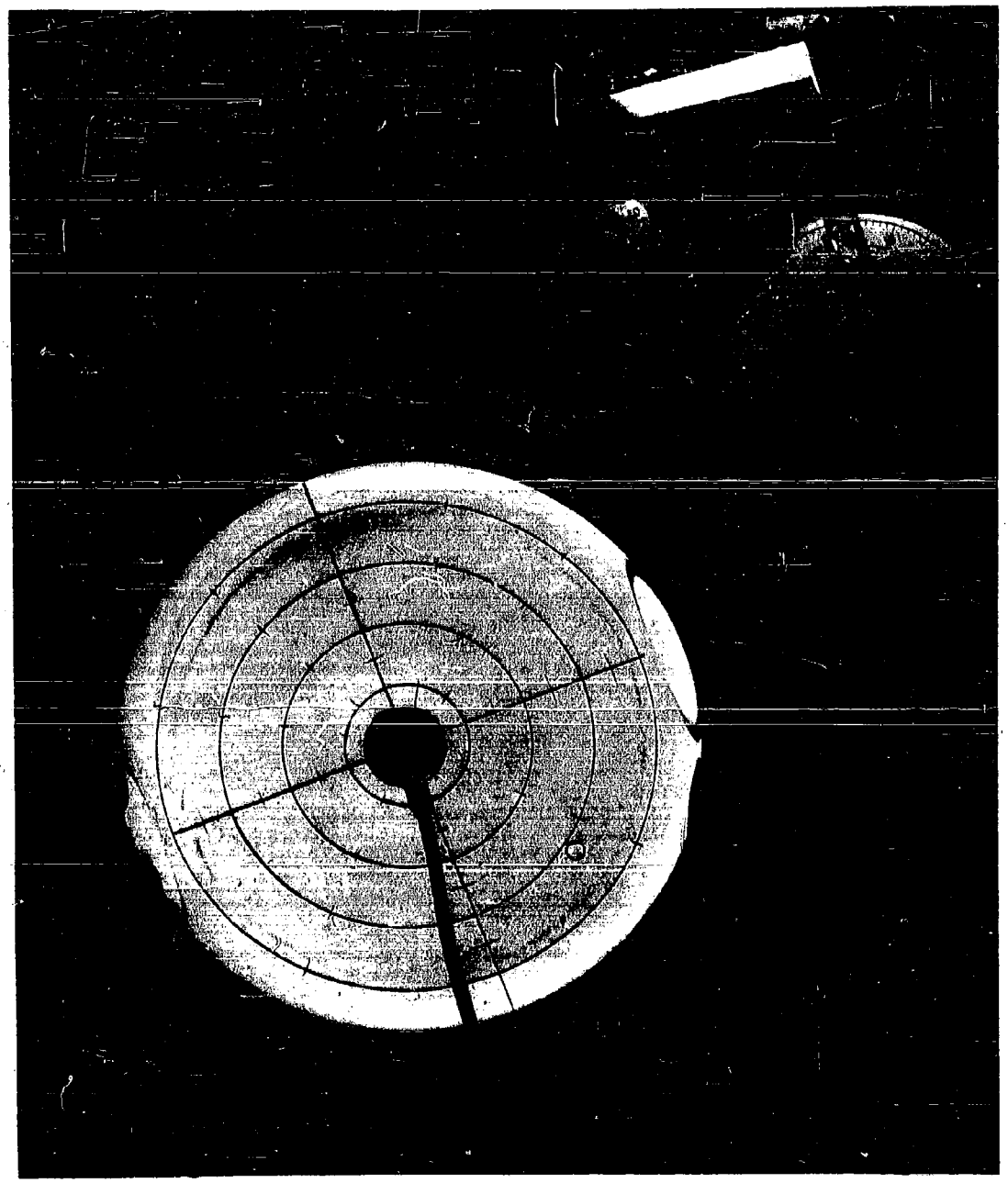


Fig. 6. Typical Droplet Collection, Initial Position

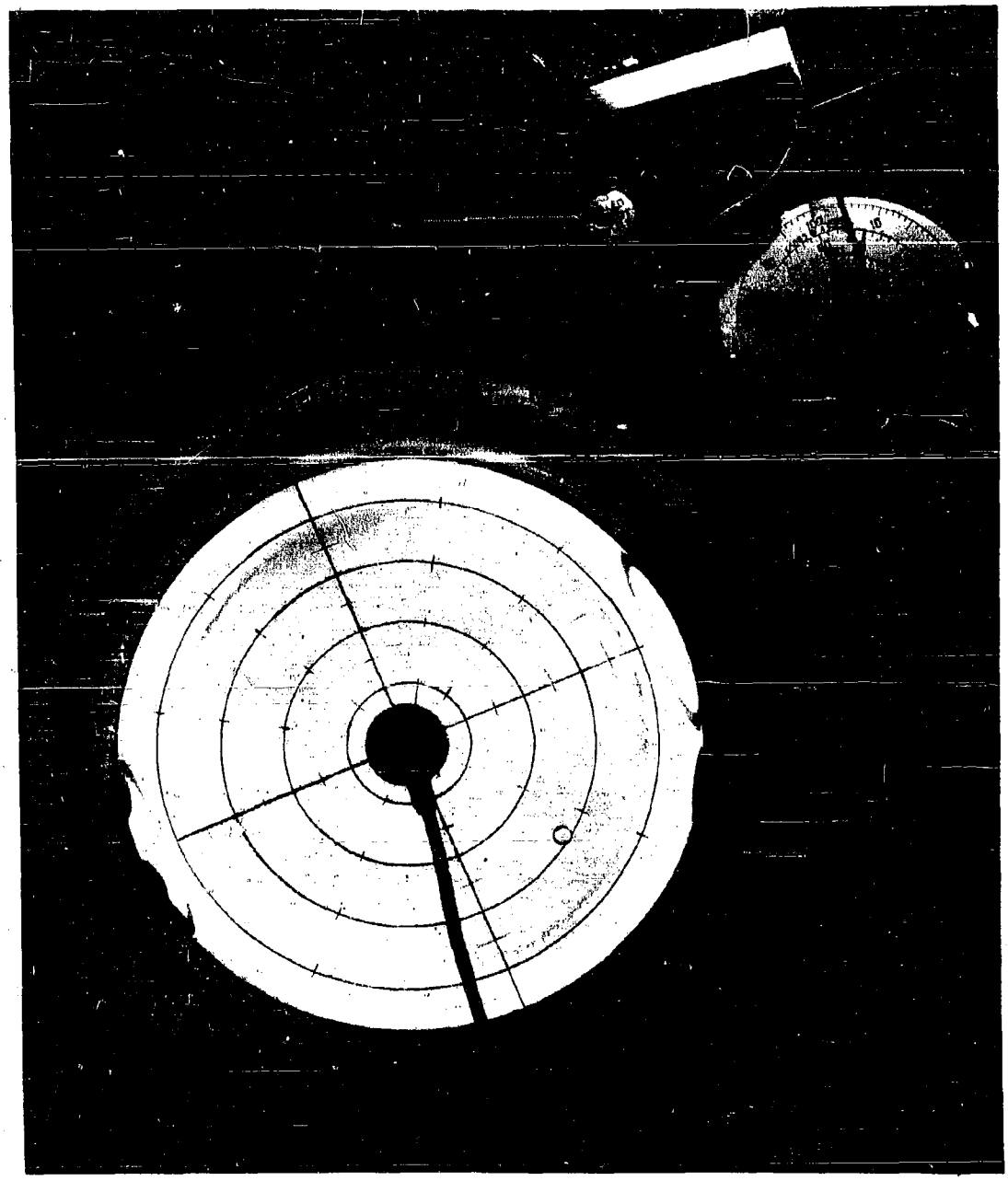


Fig. 7. Typical Droplet Collection, t = 5 Seconds

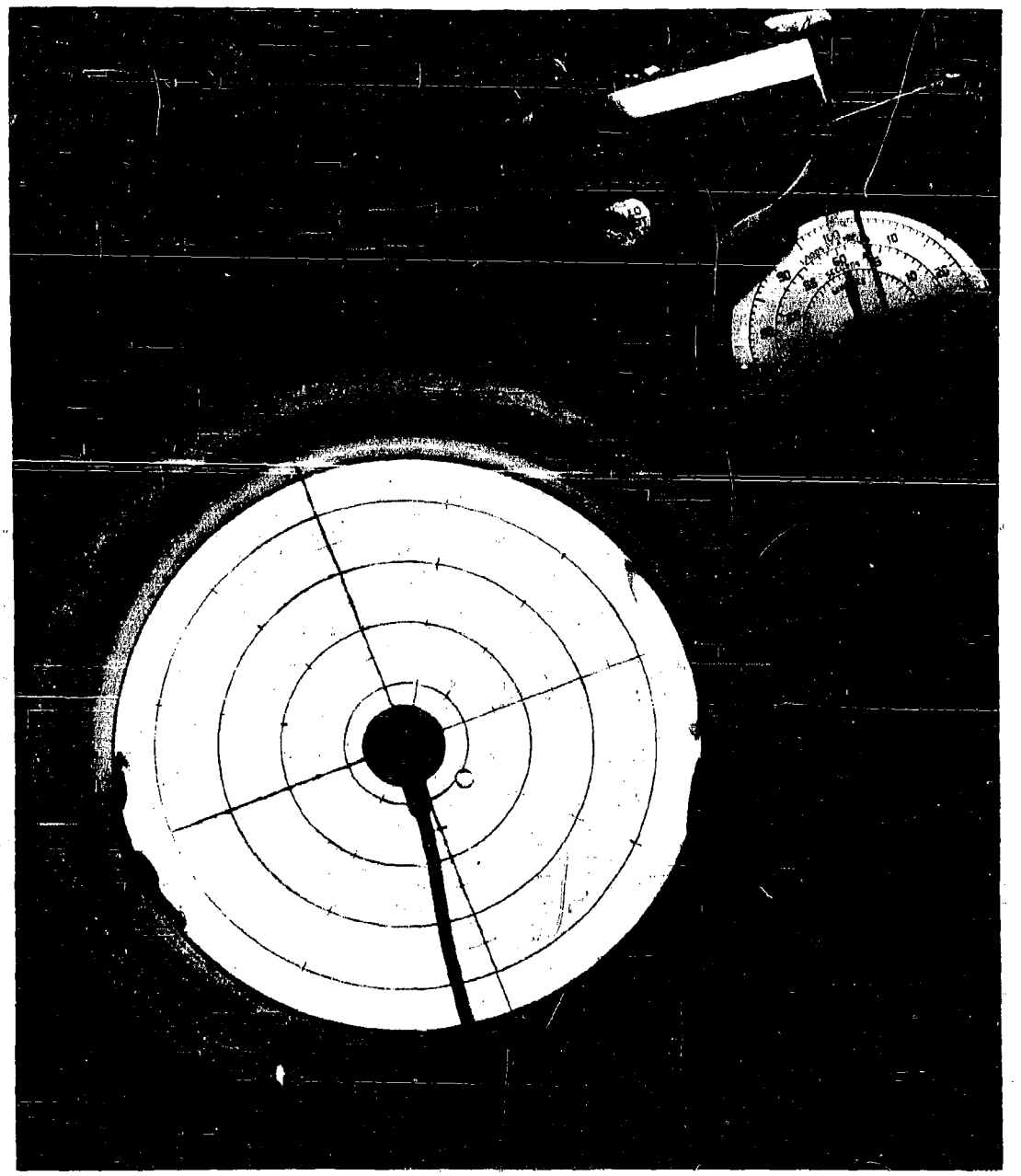


Fig. 8. Typical Droplet Collection, Prior to Contact, t = 14.2 Seconds

EXPERIMENT

Results

A representative sampling of our experimental results with the cylindrical electrode is shown in Figure 9 . This figure compares experimental data with the theoretical equation derived earlier:

$$t = \frac{3\,K''\,(\ln \frac{b}{a})^2 \eta\,(r_o^4 - r^4)}{2\,(.707)(K''-K')\,K'\,\epsilon_v\,V_a^2\,R^2} \qquad (27)$$

From the previous description of our experimental setup, it is evident that accurate measurements of the drop position are difficult to obtain. Moreover, since the radius is raised to the fourth power in the equation, any small error in measurement becomes a significant error in collection time. We feel that this accounts for most of the spread in data. Moreover, the experimental points are fairly evenly divided around the theoretical line, which indicates random experimental errors in measurement. Two other errors are present in our data. First, the drop radius was determined from the drop volume. The volume of the drop was measured by a calibrated pipette which was used to form the drop in the supporting fluid. This method of drop formation was convenient, but not particularly accurate. Another error is that drops are not perfectly spherical as is assumed in the theory. Both the density gradient and the field effect caused the drops to be slightly elliptical. However, we could not adequately measure this effect and since it was small, we decided to ignore it.

Photographs of a typical collection are shown in Figures 6 to 8 . The clock indicates the time at which the picture was taken and the position of the drop can be approximated from the photograph. The dark circular lines are one inch apart. Using the data in Figures 6 to 8 , we can calculate a sample theoretical collection time and compare it with theory.

From Figure 6 we can determine the initial droplet position, r_o, to be approximately 3.2 inches from the electrode axis. The drop volume is 0.4 milliliters and therefore R = 4.5×10^{-3} meters. The applied voltage is 7500 volts.

We know that

$$K'' = 80$$

$$K' = 2.25$$

$$\ln \frac{b}{a} = \ln 9.31$$

$$\eta = 0.00084 \ \frac{kg}{m\ sec}$$

$$\epsilon_v = 8.854 \times 10^{-12} \ \frac{farads}{meter}$$

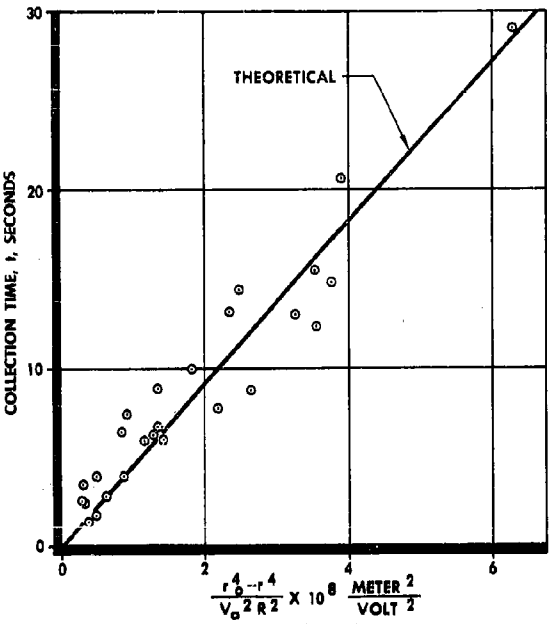


Fig. 9. Predicted and Measured Collection Times for Cylindrical Electrodes

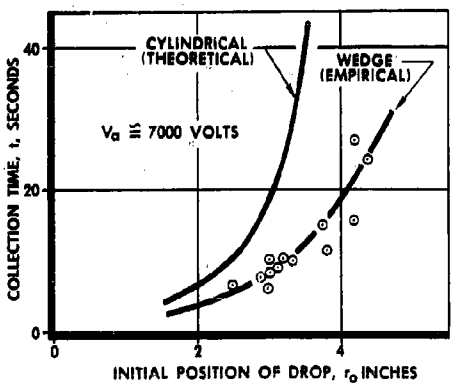


Fig. 10. Collection Times for Wedge and Cylindrical Electrodes

Substituting the above into equation (27) and solving, we have:

$$t_r = 17.2 \text{ seconds}$$

The time shown on the clock is approximately 14.2 seconds, although the droplet is approximately one inch from the rod. Our data shows the droplet contacted the rod at 14.9 seconds.

We lacked the time necessary to collect adequate data on the wedge shape electrodes, but our results are shown in Figure 10 which compares the collection time of the wedge shape electrode to the cylindrical electrodes for the same radial distance, voltage, and droplet size. It is evident that the wedge shape is more effective than the cylindrical shape. From figure 10 we can see that the wedge and cylindrical electrodes have approximately the same collection times when the drop is near the center of the tank, but the times diverge rapidly as the collection distance increases. This indicates, of course, that the wedge electrode will be more practical for large collection distances.

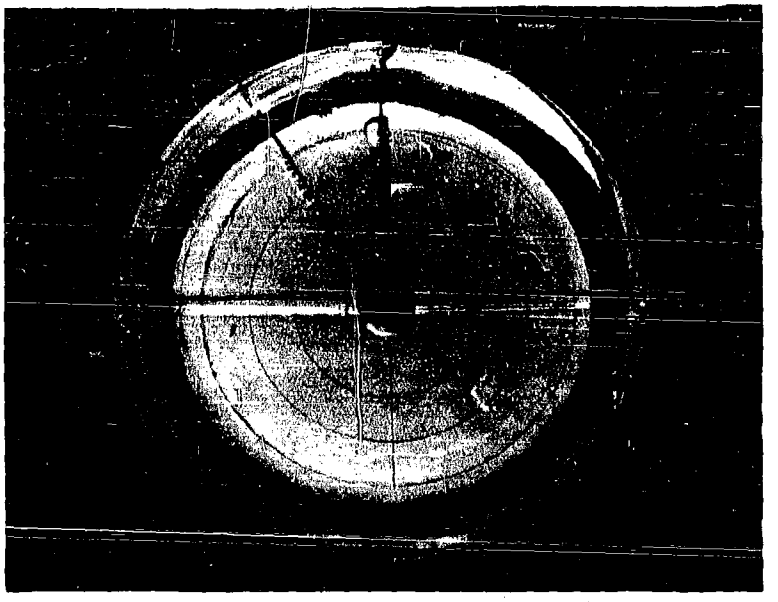
There is an additional advantage to the wedge shape as indicated by the photograph of Figure 11 . Here we see the stable configuration of a wetting fluid resulting from capillary attraction. The wedge shape electrode not only shows a shorter collection time than the cylindrical electrodes, it also has the additional collecting and stabilizing force of capillary attraction. The capillary attraction effect of a wedge with wetting fluid is discussed by W. C. Reynolds (Ref. 5).

ELECTROPHORETIC PHENOMENA

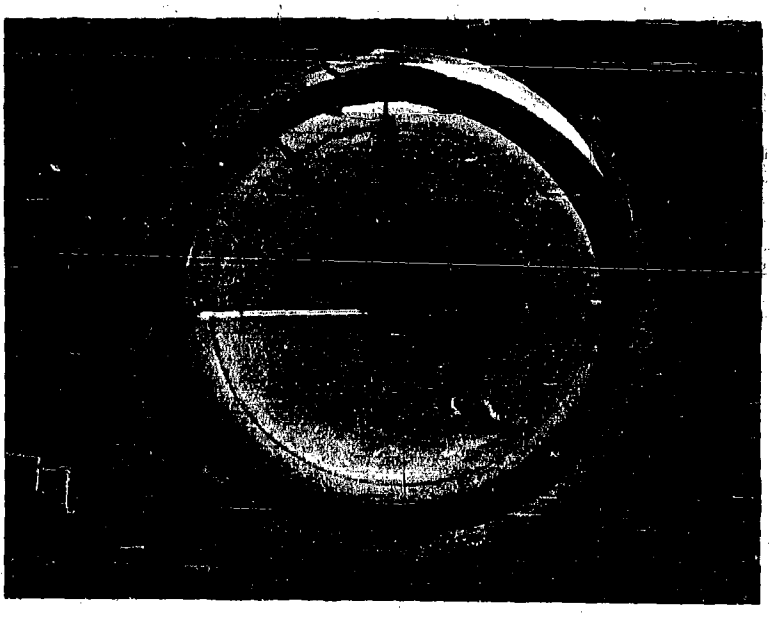
The first phase of our research was primarily concerned with dielectrophoresis, but we could not ignore the possibilities of electrophoretic effects in the control and orientation of propellants. We plan on studying the electrophoretic phenomena during our second phase of research, but we have already demonstrated several important effects which are interesting.

Electrophoresis is the motion of electrically charged bodies in an electric field; the effect is much stronger than dielectrophoresis. Some of our experiments dramatically illustrated the greater forces involved with electrophoresis. For example, electrophoretic forces were sufficient to tear water globules away from the outer wall, a virtually impossible feat with dielectrophoresis alone.

The difference between the dielectrophoretic and electrophoretic forces is easily seen. As described previously, the dielectrophoretic force is the resultant of a force of attraction and repulsion on a polarized molecule in a non-uniform electric field. The electrophoretic force is simply a Coulomb force of attraction or repulsion.



11a



11b

Fig. 11. Wedge Electrode Propellant Configuration

When our normally non-conducting supporting fluid was contaminated, it became conductive and electrophoresis would take place. In some experiments with the DC power supply, our collection times would begin to deviate drastically from the normal collection times. We found then that by reversing our electrode leads, we could alter this effect. With the cylindrical electrode model we could sometimes make the water drop move radially outward or inward, simply by reversing the electrode leads. A similar effect occurred with the wedge electrode. Evidently, the drops were collecting a net electrical charge from the current flow through the supporting mixture. By reversing the electrode leads, we reversed the current flow and hence the motion of the drop.

Another electrophoretic effect which occurred was independent of the purity of the supporting fluid. This effect was the ionization of the water drop as it came into contact with the center electrode, which has previously been described by H. Pohl (Ref. 4). At high voltages and in a region of a highly non-uniform electric field, the polarized molecules break down and become ionized. The net electrical charge of the ionized body is the same sign as that of the electrode causing the non-uniform electric field. Therefore, the body is strongly repelled. In our experiments, we found that the water drops became ionized and were repelled from the central cylindrical electrode at voltages above 2000 to 3000 volts. On occasion, we would allow a water drop to collect on the electrode and adhere because of capillary attraction. Then the voltage would be increased and the drop would break up and be strongly repelled. This effect occurred with either the AC or DC power supply, and at approximately the same voltages. We found that this effect also occurred when the central cylindrical electrode was covered by a glass test tube.

It is evident from our experiments and from Pohl's work (Ref. 4) that the breakdown effect occurs as a result of the polarizing force on the molecule. With the cylindrical electrode, this indicates that for each substance there is a critical minimum electrode radius and critical maximum applied voltage which must not be exceeded if dielectrophoretic forces alone are desired. Similarly, the collection electrodes for any dielectrophoretic system must have no regions of highly non-uniform fields because this will result in electrophoretic repulsion.

PROPELLANT COLLECTION

The dynamics of propellant collection in zero gravity are simulated, in a limited manner, by our model experiments. We have performed a few experiments with large amounts of water to simulate actual propellant collection in space. Large water globules were randomly distributed throughout the cylindrical electrode tank model to simulate a propellant tank approximately one quarter full of propellant. We could not use a higher percentage of water without incurring the risk of shorting out the electrodes. The voltage was applied and the globules collected. Because of the size and the relative amount of the globules, the assumptions used in the derivation of the theoretical equations were no longer applicable. The large globules collected rapidly because their relatively large size meant a relatively short collection

distance; in other words, the large globules did not have to move as far as small globules before contact was made with the electrode. The results of these experiments indicate that collection around a cylindrical electrode will be in the form of several separated globules until the collected mass is sufficient to merge together and form a single large globule. If the single globule is made larger by the addition of water globules, it will eventually touch the outer wall and achieve a relatively stable configuration. This stable position could, by proper design, be located at the tank outlet, as shown schematically in Figure 1c. The sequence of collection is shown in

We performed a few experiments with large globule collection on the wedge shape electrode. Such experiments illustrated that the wedge electrode has the advantages of a greater collection effectiveness and a stronger, more stable capillary attraction than the cylindrical electrode. The stable propellant configuration is shown schematically in Figure 11. The hysteresis effect of the water-to-wall contact angle is also illustrated in Figures 11a and 11b. The contact angle on the 90° wedge in Figure 11b is approximately 90°; the contact angle on the 90° wedge in Figure 11b is slightly greater than 0°. Reynolds (Ref. 5) describes the motion of a wetting fluid in the direction of convergence for a tapered capillary tube. The same effect occurs with a converging wedge, and therefore dielectrophoretic forces can be coupled with capillary attraction forces.

SYSTEM DESIGN

Since we are still in the research stage, we have insufficient information to design an optimum or even practical dielectrophoretic propellant orientation system. However, some of the design criteria and preliminary design techniques are evident. First, an orientation system must be capable of collecting propellant within a reasonable time period. Second, the propellant must be held in a stable position and located such that a constant flow to the engine is guaranteed. Third, the system must be light and have low energy requirements.

COLLECTION TIME

The electric field intensity of many different electrode shapes can be calculated from electrostatic field theory, but often such calculations are quite difficult and involved. We find that the experimental models provide a simple and economical method of comparing various complicated electrodes and determining their approximate collection time in space.

Two equations have been derived in the theoretical discussion, which give the collection time of a small propellant globule in a cylindrical tank with a cylindrical electrode. Equation (23),

$$t = \frac{2\, r_o \ln \frac{b}{a} \sqrt{r_o^2 - r^2}}{V_a \sqrt{\frac{3(.707)(K'' - K')K'\epsilon_v}{K''\rho}}}$$

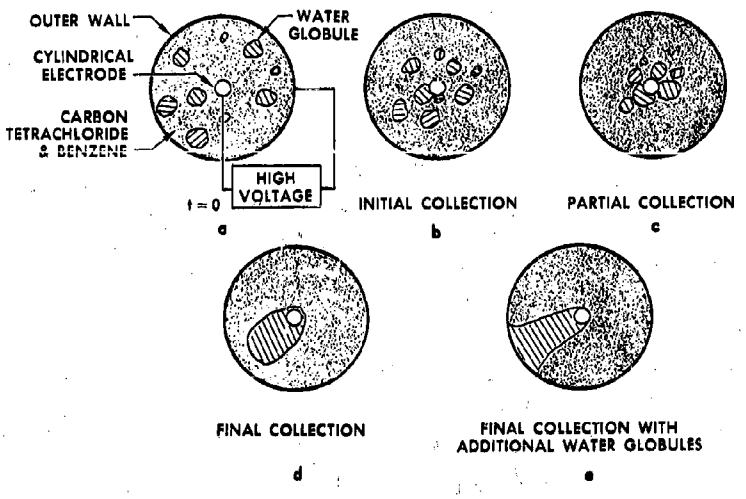


Fig. 12. Globule Collection Sequence

gives the collection time with drag neglected. This equation should hold reasonably well for the real case in space within the limits of the assumptions. The second equation, (27),

$$t = \frac{3\,K''\,(\ln \frac{b}{a})^2\,\eta\,(r_o^4 - r^4)}{2(.707)(K'' - K')K'\epsilon_v\,V_a^2\,R^2}$$

neglects the acceleration of the globule and includes drag. This equation has been adequately verified by our experiments with viscous fluids.

The collection effectiveness, or the ratio of the collection times for two electrodes, serves as a comparison for any particular electrode configuration. If we compare all electrodes with a particular cylindrical electrode, we have a measure of the relative effectiveness of each electrode desigr. In order to make adequate comparisons, the significant parameters such as voltage, collection distance, tank size, and drop size must be equal for all electrodes. After making these experimental comparisons, we may find, for example, that a particular electrode design is three times as effective as the cylindrical rod; that is, the experimental collection time for the particular electrode is one-third of that for the cylinder. Then we can determine the collection time in space. Since the particular electrode under consideration is three times as effective as the cylinder for the model experiments, then it must be three times as effective for the real case in space. In making this transition from the model experiment to the case in space, it is important to keep the electrode and tank geometries similar; all relative dimensions for the tank and the electrodes must be kept constant. Then, the theoretical collection time for the cylindrical electrode is calculated. The other electrode is three times as effective, and therefore, the collection time will be one-third of that for the cylindrical electrode in space.

As a more complete example, we can compare the collection time for the wedge and cylindrical electrodes, as is given in Figure 10. This curve shows that the wedge electrode is from two to four times as effective as the cylindrical electrode up to a distance of approximately 3.5 inches. The diverging curves indicate that the wedge electrode is far more effective than the cylinder at large initial droplet positions. Assume that we desire the collection time for a small propellant globule with the wedge electrode and tank geometrically similar to that used in the experiments. We will assume that the cylinder electrode and tank geometry is the same as that for our experiment, so that the collection effectiveness may be read directly from the curves of Figure 10. Assume that the tank in space has radii b = 2 meters, a = 0.2 meters.

The applied voltage is 50,000 volts and the propellant is liquid hydrogen with K'' = 1.27. From equation (23) we have for the collection time

$$t = \frac{2\,r_o \ln \frac{b}{a} \sqrt{r_o^2 - r^2}}{V_a \sqrt{\frac{3(.707)(K'' - K')\,K'\epsilon_v}{K''\,\rho}}}$$

Let r_o = 1 meter. Since r_o is one-half of b for the tank in space, r_o in the experimental model must be 2.5. The collection effectiveness at this point for the wedge is approximately two. From equation (23) above, we know that the collection time for the cylindrical electrode is 448 seconds. Therefore, the collection time for the wedge is 224 seconds.

There is one significant error introduced by our comparison technique. If one electrode configuration is more effective than another, then the average velocity of the collected globule is higher for one electrode than the other. From Stokes' law, the effect of drag will be greater on the faster globule, and this will decrease the relative effectiveness ratio of the electrode being considered. Electrode effectiveness ratios and predicted collection times will then be conservative.

As a preliminary design tool, the comparison technique with experimental models will be economical and simple. Moreover, effects such as capillary attraction and collection trajectories can be studied simultaneously. For these reasons, we feel the experimental models will be useful in the preliminary design of electrodes. However, final design criteria would also depend on mathematical solutions to the applicable field equations, and more sophisticated experiments in a true zero gravity environment.

FLUID PROPERTIES

From equation (23) it is evident that the collection time is a function of fluid properties. We have compared the significant parameter

$$\sqrt{\frac{(K''-K')\,K'}{K''\,\rho}}$$

for some typical propellants. The results are shown in Table 1 below. The differences in the effect of the propellant properties are quite small, considering that the range of propellant properties is large. Both the dielectric constants and densities differ by more than an order of magnitude, whereas the variation in the parameter

$$\sqrt{\frac{(K''-K')\,K'}{K''\,\rho}}$$

is only from .445 to 1.68. Therefore, the propellant collection times for various common propellant combinations will not vary by more than a factor of four.

ENERGY AND WEIGHT REQUIREMENTS

The energy requirements for the orientation systems are quite low for two reasons. First, essentially the only energy required is that necessary to charge a portion of the propellant tank to a high voltage; theoretically, there should be no current flow. Second, only the amount of propellant needed for restart will be collected. This is much more efficient than accelerating or spinning the entire vehicle in order to collect a few percent of the propellant.

Table 1

Propellants	Dielectric Constant	Specific Gravity	$\sqrt{\frac{(K''-K')K'}{K''\rho}}$
Hydrogen	1.25	.0706	1.68
Oxygen	1.55	1.14	.56
N_2O_4	2.56	1.491	.64
N_2H_4	52.9	1.011	.985
RP-1	2.08	.804	.805
Fluorine	1.43	1.52	.445

There is virtually no limit on the number of times propellant can be collected with one electrode. Therefore, the system weight increases only slightly with an increase in the number of restarts, since only the energy storage requirements are increased.

We can get an order of magnitude estimate for the energy required to orient liquid hydrogen in space by considering the following example. Assume that we have a concentric cylindrical electrode of radius a = 0.2 meters in a cylindrical tank of radius b = 2.0 meters. Let the length of the tank and electrode be L = 15 meters. For a conservative estimate, assume the tank is filled with liquid hydrogen with a dielectric constant of 1.27. Then, the capacity per unit length is given by

$$C' = \frac{C}{L} = \frac{2\pi\epsilon K}{\ln\left(\frac{b}{a}\right)} \tag{28}$$

The energy required to charge the tank to 40,000 volts is given by

$$W = \frac{1}{2} C' LV^2 \tag{29}$$

The energy required is only 0.368 joules.

The electrodes can be made quite light since they will need only sufficient structural strength to support themselves during the acceleration of the vehicle. A minimum gauge low-density metal such as aluminum would probably be sufficient. The power supply will be quite light also. For example, the power supplies used in our model experiments weighed no more than 20 pounds.

We have not attempted to make rigorous estimates of energy requirements or system weight, since the major purpose of our first phase of research is to study the physical phenomena of dielectrophoresis.

ADVANCED CONCEPTS

We have given some speculative thought to an optimum propellant orientation system utilizing electric fields. We find that the low weight and power requirements of the dielectrophoretic system make this method highly attractive. We plan to study various other electrode designs for the purpose of finding a configuration which collects propellant in a minimum time and orients it at the required location. For example, one design presently under consideration would consist of a cone with a cylindrical electrode at the axis. This design would not only collect the propellant into the center, but would also move it toward the tip of the cone, where the propellant outlet would be located.

Other designs involve more complicated shapes; for example, propellant could be collected with wedge electrodes, and passed into the axial pipe supporting the wedges. Inside the pipe, a second orientation system would move the propellant to the tank outlet location. These designs are under consideration and after further study of dielectrophoretic effects, an attempt will be made to develop a practical system.

Propellant collection times in large tanks could be prohibitive, even with expected improvements in electrode design. The average collection time could be decreased by using several small electrodes instead of one large, centrally located electrode. The smaller electrodes would have larger field gradients and the propellant collection distances would be decreased. However, this method would increase the system weight.

Therefore, in our future research, we will also be considering the application of electrophoretic effects to the collection of propellant in zero gravity. As discussed previously, the electrophoretic forces are significantly larger than dielectrophoretic forces; the collection times will therefore be significantly less than with dielectrophoresis.

Electrophoresis offers other possible advantages. First, it may be useful for moving wall-bound propellant prior to restart. Such a system would be required if all of the propellant were located away from the tank outlet in a stable equilibrium position against the tank wall. Second, it may offer a practical means of separating liquid from gas in a gas pressurization vent system. Third, it may provide a means of stirring propellant in zero gravity; this would aid in propellant heat transfer problems by decreasing localized boiloff due to high heat fluxes.

However, electrophoresis requires much more power than dielectrophoresis and for this reason a compromise system utilizing both effects may be the best method of orienting propellants in zero gravity.

CONCLUSIONS

Although the dielectrophoretic propellant orientation system has only recently entered the experimental stage, results have been encouraging. The physical effect has been amply demonstrated and theoretical calculations

have been experimentally verified. The system appears to have significant advantages. The power requirements should be low and the system weight required is apparently small. Modifications such as electrophoretic orientation and the auxiliary use of capillary action make the system appear more practical. Good preliminary design criteria should be easily obtained with the experimental models. As an alternative approach to propellant control in zero gravity, dielectrophoresis should be considered a strong possibility.

NOMENCLATURE

A	Cross-sectional area of rod
a	Radius of axial high-voltage rod
b	Inner radius of tank
D_n	Electric displacement vector
E	Electric field intensity
F	Stress or force per unit area
K	ϵ/ϵ_v
R	Radius of globule
r_o	Initial radial distance of globule from center of rod
r	Radial distance of globule from center of rod
V	Volume
V_a	Voltage of rod
v	Velocity of globule
ϵ	Capacitivity
ϵ_v	Capacitivity in vacuum
ρ	Density of globule
η	Coefficient of viscosity

REFERENCES

1. Unterberg, W. and Corgelliere, J., "Zero Gravity Problems in Space Power Plants," ARS Journal, June 1962, 32 (6): 862-872.

2. Smythe, William K., "Static and Dynamic Electricity," McGraw-Hill Book Company, 1950.

3. Frank, Nathaniel H., "Introduction to Electricity and Optics," McGraw-Hill Book Company, 1950.

4. Pohl, Herbert A., "Non-Uniform Electric Fields," Scientific American, December 1960.

5. Reynolds, William C., "Hydrodynamic Considerations for the Design of Systems for Very Low Gravity Environments," Stanford University, September 1, 1961, Technical Report Number IG-1.

THE PREDICTION OF HEAT TRANSFER AND ABLATION IN THE AFT-CLOSURE OF A SOLID PROPELLANT ROCKET MOTOR

W. C. Kuby, Jr.
J. L. Richardson

Ford Motor Company
Aeronutronic Division
Research Laboratory
Newport Beach, California

ABSTRACT

A mathematical design model has been developed for the prediction of insulation requirements in the aft-closure of a solid propellant rocket motor. This model and the associated calculation technique represent only a first approximation to the exact solution of a problem concerned with various complex phenomena. The behavior of ablative elastomeric insulation materials in such an application is considered. The choice of the assumptions used to calculate the forced convective and radiative heat fluxes is based on recent experimental results. A summary is made of the information still to be determined in order that this model can be fully utilized.

The work described in this paper was performed under Air Force Contract No. AF 04(611)-7047, sponsored by the Air Force Flight Test Center, Air Force Systems Command, Edwards Air Force Base, California.

THE PREDICTION OF HEAT TRANSFER AND ABLATION IN THE AFT-CLOSURE OF A SOLID PROPELLANT ROCKET MOTOR

W. C. Kuby and J. L. Richardson
Ford Motor Company
Aeronutronic Division Research Laboratory
Newport Beach, California

INTRODUCTION

A continuing problem in the field of rocketry has been the protection of the rocket walls from the extreme temperatures and the high heat fluxes that exist. This problem has become more difficult in the case of solid propellant rocket motors since the introduction of metal powders to increase performance. The existence of the condensed-phase metal oxides in the combustion effluent has proved to create a severe environment, particularly in the aft-closure region. Despite the importance of understanding the phenomena that lead to this problem, no dependable method has been developed to predict the deleterious results of motor firings. The purpose of this paper is to delineate the problem and to indicate the areas in which more research work is necessary in order to understand the complex phenomena sufficiently to analytically predict the consequences in an unfilled (or filled) aft-closure region. In order to accomplish this end a calculation technique is outlined, and where suitable analyses exist, these analyses are discussed in detail. In those areas in which suitable analyses do not exist, the problem is defined and the necessary understanding indicated. Due to the complexity of the problem and the dearth of information in certain areas, many effects have not been considered. Among these effects are the complications of boundary layer reactions and their effect on heat transfer rates and surface regression rates, and the reduction of convective heat transfer caused by blowing.

DISCUSSION OF MODEL

The discussion presented herein will apply to an aft-closure configuration such as is shown schematically (see Figure 1). Although this axisymmetric case is possibly the simplest of configurations that one might choose, it merely serves as a starting point for such discussions of aft-closure problems. Theoretical analyses of subsonic jets of compressible fluids flowing into a compressible static fluid as discussed by Pai (1) show that the spreading angle of the jet is very small. These results have been verified by water model studies by Price, et al. (2), and they suggest that there are two separate flow regions in the aft-closure, one of which is a recirculation zone existing above the intersection of the boundary edge of the jet and the aft-closure wall; the second is the fast flowing fluid below this intersection. The significance of such a model is that there exist two distinct zones in which the factors controlling the ablation processes may be completely different. The first, the recirculation zone, is characterized by low convective heat transfer and

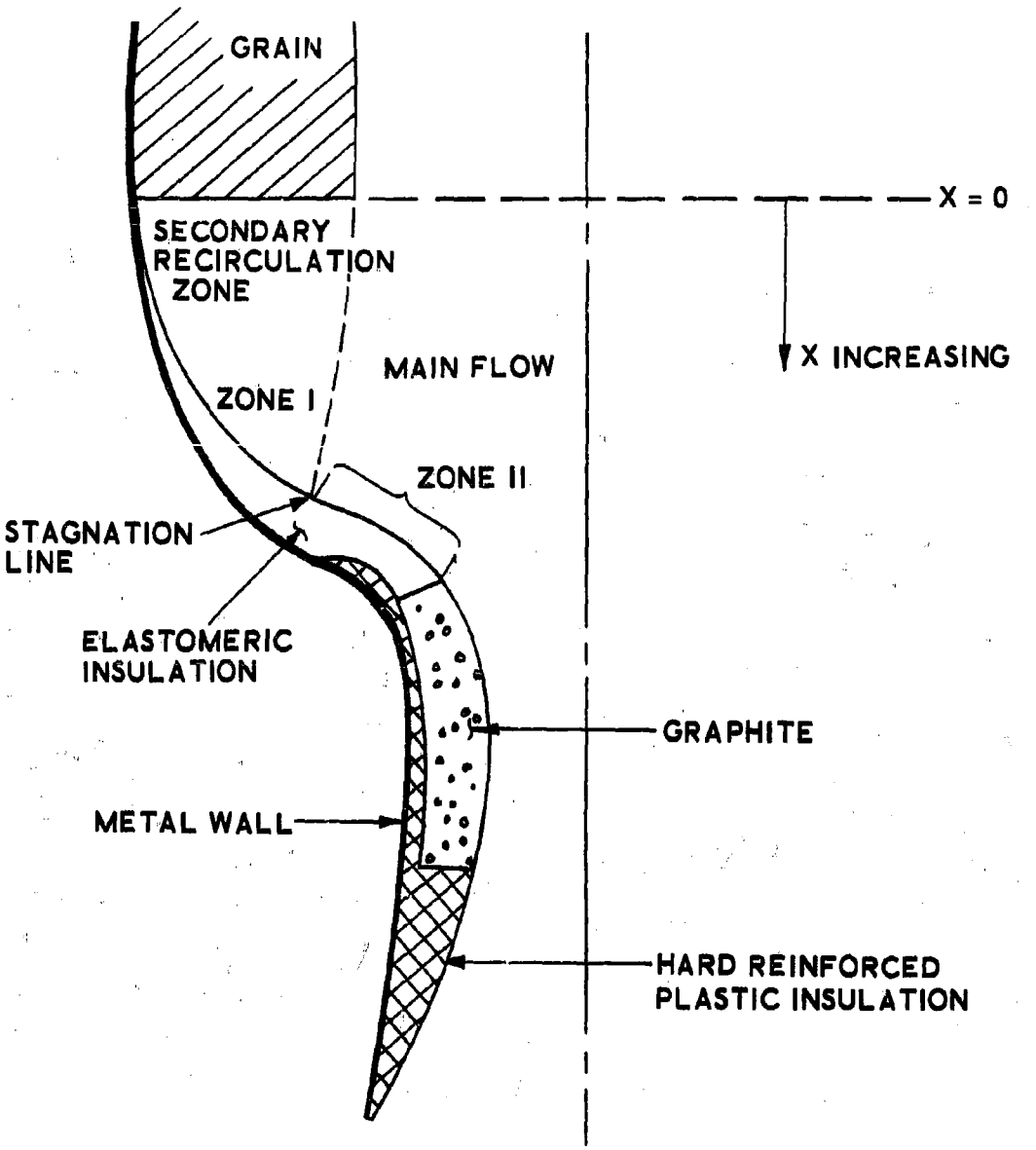


Fig. 1. Configuration at a Given Time

negligible particle effects. If the materials used in this zone form a char layer, the char layer growth is the controlling factor and the heat flux need not be determined. For non-charring materials the regression rate is dependent on the amount of heat flux to the surface. In the second region, the ablation rate is controlled by the amount of heat supplied to the surface and by the erosion of the particle stream. Another implication of this concept of two flow zones is that the particles contained in the outer regions of the jet must necessarily go through a severe turn in the general region of the aft-closure wall. This turn gives rise to high centrifugal forces which cause these particles to "slip" from the flow and impinge upon the aft-closure wall. The result of such impingement is an increase in the amount of heat transfer to the wall as well as an increase in the amount of mechanical erosion of the insulation materials. Finally, since the grain port hydraulic radius is a function of time, the extent of the two zones is a function of time. Consequently, the calculation technique should account for this time dependence. Also because of the movement of the separation point, the boundary layer, and thus the convective heat flux in Zone II, have time variations.

If use is made of the above-stated model having two flow regions in the aft-closure (see Figure 1), the problem necessarily separates into two approaches. In the recirculation zone for char-forming insulators, there exists a transient ablation process which is controlled by the char layer growth rather than the heat and/or erosion factors associated with the gas-particle stream. An analysis of this problem has been made by Grosh (3). The non-char case must be handled by empirical methods. In the second region the analysis to be used is based on the independent evaluation of the heat fluxes and the assumption that the total heat is the additive effect of each individual mode of heat transfer. Methods are discussed for calculating the convective heat flux as well as the radiative heat flux. Evaluating the heat flux associated with particle impingement as well as determining the mechanical erosion is found to be very difficult with the present state of the knowledge. In order to relate the heat fluxes and mechanical erosion to the aft-closure insulation material, a concept similar to the currently used Q* (see Eq. 38) is suggested. The main difference is that the parameter used in this paper is tailored to the requirements of the analysis discussed herein. Although the techniques put forth in this paper do not allow an accurate and complete determination of the consequences of the gas particle flow in a solid propellant motor aft-closure, they do indicate what knowledge is necessary to successfully attack the problem that exists.

ZONE I: THE SECONDARY FLOW REGION

Char Forming Materials - The Grosh Analysis*

In the secondary recirculation region it is assumed that there are very low surface-regression rates, and as a result the primary consideration is the transient transfer of heat through the insulation material with or without the formation of a surface char layer. The former case has been analyzed by R. J. Grosh (3), and Scala and Gilbert (4). Grosh's equations have been programmed for use in this region.

*The nomenclature for this section is defined as it occurs in the text.

The Grosh analysis is based on a simple extension of the well-known Neumann problem (5). For the case of a single phase change (i.e., virgin ablative material pyrolyzing to form a porous char and gaseous products), Grosh used the following one-dimensional equations to describe the transient temperature distribution in both the porous char layer and the virgin ablator (see Figure 2):

$$\frac{\partial T}{\partial t} - \nu\left(\frac{dX_1}{dt}\right)\left(\frac{\partial T}{\partial x}\right) = \alpha_1 \frac{\partial^2 T}{\partial x^2} \qquad 0 < x < X_1 \tag{1}$$

$$\frac{\partial T}{\partial t} = \alpha_s \frac{\partial^2 T}{\partial x^2} \qquad X_1 < x < \infty \tag{2}$$

where

T = temperature at any point (°F)
t = time (sec)
x = position measured relative to the stationary exposed outer edge of the char layer (ft)
α_1 = thermal diffusivity of the gas-porous char layer (ft^2/sec)
α_s = thermal diffusivity of the virgin ablative material (ft^2/sec)
X_1 = time-dependent position of the inside edge of the porous char layer (ft)
ν = parameter proportional to the difference in density of the gas in the porous char layer and the virgin ablative material (defined in Reference 3)(dimensionless)

and the initial and boundary conditions used are:

$$T = T_i \qquad x = x,\ t < 0 \tag{3}$$

$$X_1 = 0 \qquad t < 0 \tag{4}$$

$$T = T_o \qquad x = 0,\ t \geq 0 \tag{5}$$

$$T = T_1 \qquad x = X_1,\ t \geq 0 \tag{6}$$

$$T = T_i \qquad x = \infty,\ t = t \tag{7}$$

and

$$\left(\frac{\partial T}{\partial x}\right)_t \Bigg|_{x=X_1^-} = \gamma_1 \left(\frac{\partial T}{\partial x}\right)_t \Bigg|_{x=X_1^+} + \omega\left(\frac{dX_1}{dt}\right) \tag{8}$$

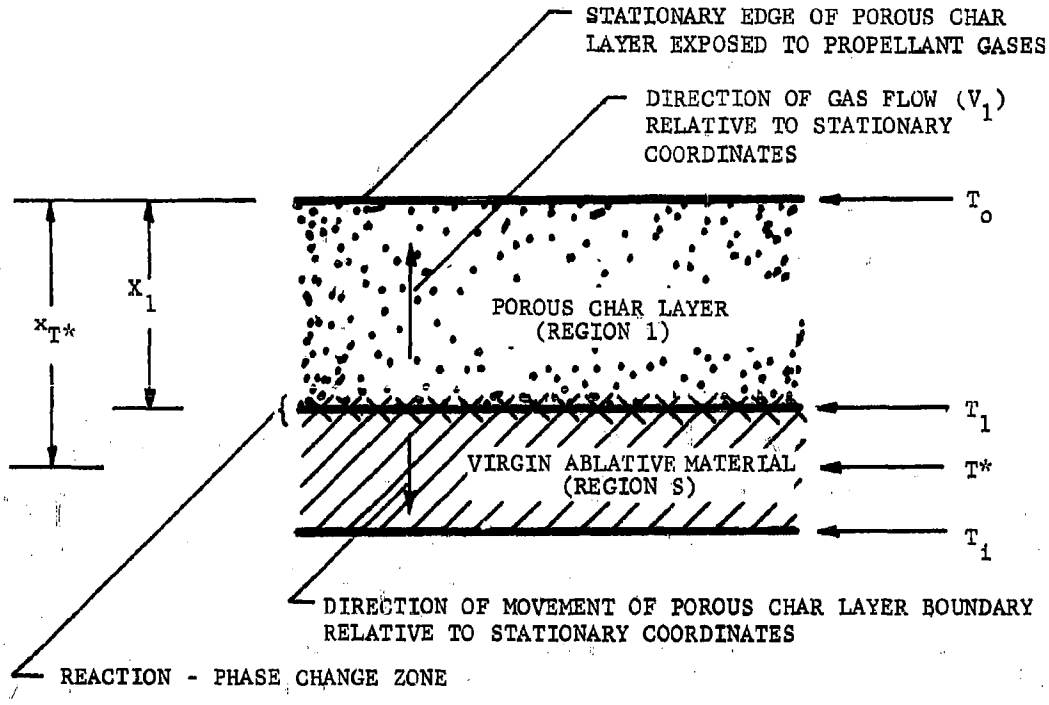


Fig. 2. One-Dimensional Char-Forming Ablator - Char Layer Not Removed (Symbols Apply to Grosh Analysis)

where γ_1 = the ratio of the thermal conductivity of the virgin material to that of the gas-char composite

and ω = quantity proportional to the sum of the heat of reaction (pyrolysis) and heat of sublimation (or heat of melting plus heat of vaporization) (sec °F/ft^2).

For this one-dimensional situation several important assumptions have been made in addition to those implied by Eqs. (3) through (8):

(a) The gas moves or transpires so slowly through the porous char that these two materials are in thermal equilibrium throughout the entire char layer.

(b) There is a step change in the physical properties of the char and virgin material at $x = X_1$, and the reaction zone thickness (i.e., presence) can be neglected.

(c) The physical properties of the materials are not temperature-dependent (average values employed).

(d) The composite (gas-char) thermal conductivity and mass-density/heat-capacity product can be adequately described by linear, porosity-additive expressions utilizing the properties of the pure materials.

The solution to the above equations is found in a manner similar to that utilized to obtain a solution to a modified Neumann problem (see Reference 5, p. 291, Eqs. (51) and (52)). The char depth is determined by the relation

$$X_1 = bt^{1/2}, \tag{9}$$

where b is the eigenvalue solution of a transcendental equation obtained from the moving-boundary condition (Eq.(8)) and is a function of the properties of the materials involved and the boundary temperatures assumed. For insulation materials which might be expected to conform to the Grosh model, it should be possible to measure b directly in the laboratory under the same conditions of surface temperature and pressure as would be encountered in the aft-closure which is being studied. The b so obtained could then be compared to that computed theoretically in order to arrive at better estimates for the char and virgin material properties.

The ablative materials for aft-closure applications which appear to be of greatest current interest are elastomers. Some form a char layer during ablation, while others form very little char, and that which does result has especially weak structural characteristics. Presently little is known about the properties of the char formed from elastomeric materials.

Using an elastomeric material and making estimates of the properties of the char produced by it, the computer program solution of the Grosh equations* is utilized to obtain the char thickness (X_1) and the heat penetration distance (x_{T*}) as a function of time. The x_{T*} is obtained from the following equation:

$$\left(\frac{T^* - T_i}{T_1 - T_i}\right) = \frac{\operatorname{erfc}\left\{\frac{x_{T*}}{2\sqrt{\alpha_s t}}\right\}}{\operatorname{erfc}\left\{\frac{b}{2\sqrt{\alpha_s}}\right\}} \tag{10}$$

(erfc denotes the complementary error function), where x_{T*} is the position at which the temperature is T* for a given time. The properties and boundary temperatures assumed for the example cases presented here are given (with their sources) in Table 1.

The results are presented in Figures 3 and 4. The calculations were performed for several different assumed values of the unknown properties of an elastomeric material. Cases 2, 3 and 4 utilized identical input information as employed in case 1 except for the one key property designated on the figures. For example, for case 3 the three relatively unknown properties which were assumed are the same as those used in case 1 except that L_1 = 200 Btu/lb is used in place of L_1 = 500 Btu/lb. The following is evident:

(1) The thermal conductivity of hydrogen is of greatest importance in determining the overall thermal conductivity of the transpiring gas phase, $k_{f,1}$ (for $x_{H_2} \geq 0.1$, say).

(2) The heat-penetration distance, x_{T*}, is only slightly greater than the char-layer depth, X_1, because of the much lower thermal conductivity of the virgin elastomeric material compared with that of the porous char (note that $k_{m,1}/k_s$ is on the order of 30 for $k_{m,1} \sim 1 \times 10^{-3}$ Btu/ft sec °F).

(3) For assumed values of gas composition and boundary temperatures, the results are most sensitive to the value assumed for $k_{m,1}$ (thermal conductivity of solid matrix material comprising the structure of the porous char). Considering the discussion given in Note 7, Table 1, concerning the range of thermal conductivity values which might be realistic estimates for the physical-thermal situation of interest here, and considering the fact that the thermal conductivity

*It should be noted that several typographical errors appeared in the original Grosh analysis (3). These were corrected before the equations were programed for the computer.

Table 1. Input to the Grosh Analysis

Properties and Boundary Temperatures	Explanations and Sources
1. T_o = 3000 °F	1. Based on measurements of reference 39.
2. T_1 = 650 °F	2. Estimate based on Aerojet material test data reference 7.
3. T_i = 100 F	3. Ambient conditions.
4. σ = 0.40, 0.65	4. Range of porosities typical of plastic chars (see reference 38).
5. (a) L_{1max} = 500 Btu/lb, (b) L_{1min} = 200 Btu/lb	5. Based on a rough extrapolation of Q* data to zero heat flux (reference 41) and estimates given for plastic ablators (see reference 38).
6. $k_{f,1} = 8 \times 10^{-5}$ Btu/ft sec °F	6. Based on assumed gas composition at 1 atm pressure and 1500°F; linear additivity combining rule used; data from reference 6, p. 461; pressure effect neglected; approximate extrapolation of data from 600 to 1500°F.
7. $k_{m,1}$ = (a) 1×10^{-3} Btu/ft sec °F, or (b) 8×10^{-3} Btu/ft sec °F	7. Data for non-porous petroleum coke at 200°F given on p. 457, reference 6. Data for σ = 0.40 porous graphite at 1500°F (see reference 42). Note that at 1000°F the thermal conductivity of this material is about 0.8×10^{-3} Btu/ft sec °F and that petroleum coke of 20-100 mesh has a thermal conductivity of 0.15×10^{-3} Btu/ft sec °F at 800°F, while powdered coke has a thermal conductivity of 0.031×10^{-3} Btu/ft sec °F at 200°F.
8. $k_s = 3.78 \times 10^{-5}$ Btu/ft sec °F	8. Thermal conductivity at 300°F, reference 41.
9. $\rho_{f,1}$ = 0.55 lb/ft^3	9. Computed from the perfect gas mixture law: $\rho_{f,1} = \frac{P_1 M}{RT_1}$, where $M = \sum_{i=1}^{4} x_i M_i$ (R = gas constant = 10.731 $\frac{ft^3\ psia}{lb\ mole\ °R}$; M_i = molecular weight of species i) with P_1 at 500 psia, T_1 at 1500°F.
10. $\rho_{m,1}$ = 94 lb/ft^3	10. Density of carbon stock (see reference 6, p. 457).
11. ρ_s = 80.5 lb/ft^3	11. Density at 300°F, reference 41.
12. $C_{f,1}$ = 0.42 Btu/lb °F	12. Frozen isobaric specific heat for assumed gas mixture composition $C \equiv C_p = \frac{\sum_{i=1}^{4} x_i \tilde{C}_{p_i}}{M}$, where $\tilde{C}_{p_i}$ = isobaric molar heat capacity of species i obtained from average values (60 to 3000°F) reported in reference 43 (M = 23 lb/lbmole).
13. $C_{m,1}$ = 0.387 Btu/lb °F	13. Specific heat for carbon. Average value for temperature range 130 to 2640°F (see reference 6, p. 235).
14. C_s = 0.48 Btu/lb °F	14. Heat capacity at 300 F, reference 41.
15. T* = 200 °F	

Definitions of Terms

x_i = mole fraction of species i

σ = porosity of char

L_1 = heat of reaction + heat of phase change

k = thermal conductivity (Btu/ft sec °F)

ρ = mass density (lb/ft^3)

C = isobaric specific heat (Btu/lb °F)

Subscripts

1 denotes region 1, porous char-gas layer

s denotes region s, virgin material

f denotes gas evolved

m denotes solid matrix material of porous char

$$\left(\quad \right)_{f,s} \equiv \left(\quad \right)_s$$

Assumed Composition of Gas Given Off by the Ablating Elastomeric Material; $x_{H_2} = x_{H_2O} = x_{CO_2} = x_{CO} = 1/4$

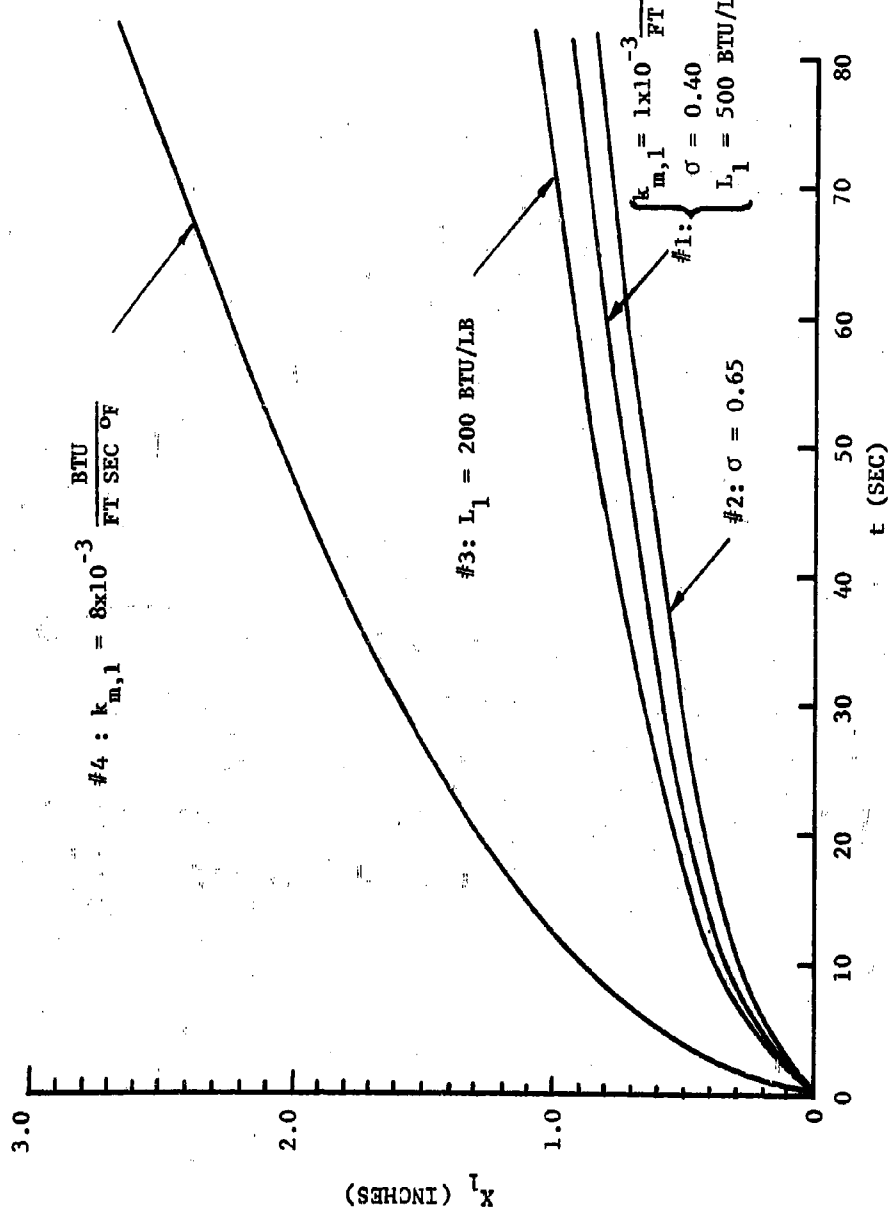


Fig. 3. Char Layer Thickness as a Function of Time

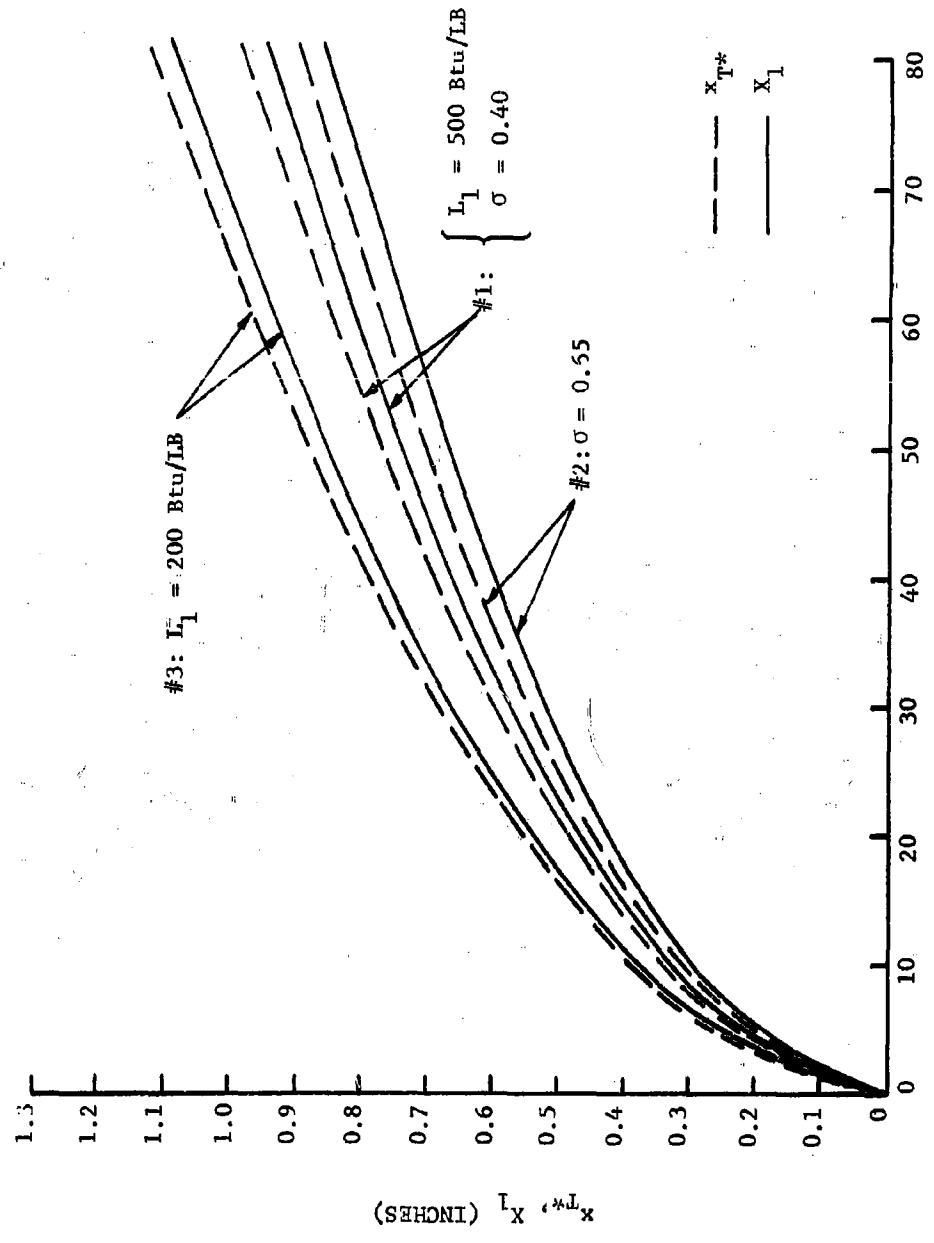


Fig. 4. Original Insulation Thickness Required to Keep Back Wall at 200°F and Char Layer Thickness as a Function of Time

of these materials in many cases decreases with increasing temperature, it is possible to assume that the $k_{m,1}$ should be somewhat less than 1×10^{-3} Btu/ft sec °F. (For example, the room temperature thermal conductivity of a porous carbon such as Carbocell grade 10 has $k_{m,1} \sim 0.5 \times 10^{-3}$ Btu/ft sec °F at room temperature (see Reference 6, p.1550).) This would yield char thicknesses on the order of tenths of an inch or less — a result which is compatible with char thicknesses actually obtained in hot test firings (7). The accuracy of the $k_{m,1}$ estimate will be the most important factor which will limit the usefulness of the Grosh analysis — or any other theoretical attempt to predict char-layer growth.

(4) The value assumed for L_1, the overall heat sink term (reaction plus phase change) has a modest influence on the results for the range of L_1 values which seem to be most appropriate for these elastomeric materials.

(5) For $0.4 < \sigma < 0.65$, porosity has little influence on the results.* This would not be expected to be true for extreme values of the porosity, i.e., for $\sigma \gg 0.65$ or $\sigma \ll 0.4$.

Nonchar-Forming Insulation Materials

If no char forms during the ablation process, the rate of material removal will be controlled by the heat flux to the surface. This heat flux is assumed to be the sum of convective and radiative fluxes. In a subsequent section of this paper a method is given for determining the convective flux in Zone II and the radiative flux in the aft-closure region. Since the velocities in the recirculation zone may be of the order of one-third those in the free stream (8), the convective heat transfer coefficient will also be about one-third the free stream value (see also Reference 9). Considering a gas-particle flow at 5500°R, a surface temperature of 3000°R, and a Zone II convective heat transfer coefficient of 0.5×10^{-3} Btu/in.^{2} °F sec, the convective flux is about 10% of the total flux. As a first approximation, therefore, one can consider only the radiative flux.

By using laboratory test data for Q*, the effective heat of ablation, the linear regression rate for the material is obtained from the following:

$$r \sim \frac{q_R}{\rho_{in.} Q^*} \tag{11}$$

*Recent laboratory measurements of the porosity of an elastomeric char indicated that σ (volume porosity measured at ambient conditions) at the top of the char layer which was exposed to the propellant gases was about 0.66, and at the reaction zone — char layer interface, was 0.75. The density of the solid matrix material, $\rho_{m,1}$, was about 103 lb/ft^3 (44).

This approximate value of the rate will be constant for any axial position in the aft-closure. An upper limit on the value of the regression rate is obtained by adding in a convective flux based on the forced convective transfer coefficient calculated for the first portion of Zone II.

ZONE II - THE HIGH HEAT FLUX REGION

In this region the controlling factor is the heat flux to the surface. This flux is assumed to be composed of linearly additive contributions from the mechanisms of forced convective, radiation, and particle impingement with energy accommodation at the surface.

Calculation of Convective Heat Transfer

Selection of Method. Analysis of aft-closure heating requires a method of predicting convective heat-transfer coefficients. Experimental results (10), (11) show that the convective heating in a rocket motor may vary by a factor of two or more, depending to a large extent on combustion variables. For this approximate analysis of aft-closure heat transfer, the effects of wall roughness, combustion instability, chemical reactions in the turbulent boundary layer and blowing have been neglected.* Each of the first three of these phenomena could be expected to increase the convective heat transfer rate to the surface, while the latter phenomenon would be expected to decrease the net heat flux to the wall.

The simplified Bartz (10) equation

$$h_c = \left\{ \frac{0.026}{D_*^{0.2}} \left(\frac{\mu^{0.2} C_p}{N_{Pr}^{0.6}} \right)_0 \left(\frac{P_c g}{C^*} \right)^{0.8} \left(\frac{D_*}{r_c} \right)^{0.1} \right\} \left(\frac{A_*}{A} \right)^{0.9} \sigma \qquad (12)$$

has been widely used for predicting heat transfer from rocket exhausts and is certainly accurate enough to do this within the scatter of the experimental data. Nevertheless, there is an important reason for using a more sophisticated method than the simplified Bartz equation for this analysis. One of the more important considerations in the aft-closure study is the prediction of the interaction between the aft-closure geometry and the heat transfer. Since the simplified Bartz equation is based on an empirical one-dimensional pipe flow equation assuming fully developed turbulent flow, it does not readily allow the inclusion of the effects of local variations in velocity and static pressure associated with the flow geometry. Instead, a method is required that will show these local time-varying effects of the nonuniform flow field in the aft-closure region upon the boundary layer and consequently the heat-transfer coefficient.

---

*Under conditions of chemical equilibrium in the propellant gas phase, the effects on the convective heat transfer due to chemical reactions can be appreciable in relation to the other effects which have been neglected (12), (13), and (14).

The Culick and Hill (15) application of the Truckenbrodt turbulent boundary-layer analysis (16) was selected. The Truckenbrodt analysis was chosen because of (a) its proved accuracy in incompressible boundary-layer analysis, (b) the ease with which time-dependent flow behavior can be accounted for, and (c) because, by using velocities at the edge of the boundary layer as input data (obtained from modeling, say, or a one-dimensional isentropic compressible flow calculation), it can include the effects of circumferential variations in the flow field in the estimation of the skin-friction coefficient. This analysis is recommended and discussed in detail by Schlichting (see Reference 17, p. 580), and is valid for any arbitrary pressure gradient and velocity change at the edge of the boundary layer. It is based on the work of Rotta (17),(18), and depends on the stepwise integration of the energy integral equation along the surface. Culick and Hill have shown that, provided

$$H + 1 = (H_i + 1)\ (T_\infty/T) \tag{13}$$

where H is the boundary-layer shape parameter (see Equation (15) of this report and Figure 22.7 of reference 17), it is possible to apply a Stewartson-Illingworth transformation to the y coordinate and convert the incompressible Truckenbrodt analysis to one valid for compressible flow. Culick and Hill have shown that for the case of little or no neat transfer, Equation (13) is valid for Mach numbers less than 5.

As an initial condition, this method requires a knowledge of the conditions of the laminar boundary layer just before transition to turbulent flow (ordinarily obtained from a laminar boundary-layer analysis). However, it should be sufficient for most aft-closure designs to assume a zero pressure gradient at transition and an arbitrarily thin momentum layer thickness, or to estimate the conditions at transition by considering a laminar flat-plate boundary layer using Michel's (19), (20) method to locate transition. It is not expected that the heat transfer coefficients calculated will be sensitive to the initial conditions, except very close to the transition point (see discussion relative to Figure 5).

Program Formulation. The Culick and Hill modification involves the use of several transformations to convert compressible flow input parameters into equivalent incompressible values. These equivalent incompressible flow quantities are used in the Truckenbrodt method which consists of the following steps: (1) the determination of the incompressible momentum thickness by an integration of a function of the velocity profile, (2) an integration involving the incompressible momentum thickness, θ_i, and a function of the velocity profile to determine the incompressible boundary layer shape factor L_i, and (3), the determination of the related shape factor H_i. Knowledge of θ_i and H_i as a function of axial position permits the calculation of the incompressible skin friction coefficient, C_{f_i}. The resultant C_{f_i}'s are converted to the equivalent compressible values of C_f. Finally, by using a modified Reynolds analogy (nearly identical in form to the Chilton-Colburn modification) the forced convective heat transfer coefficients are calculated from the values of C_f. Of course, in general the entire calculation must be repeated for several times during the firing in order to ascertain an adequate description of the time variation of h_c.

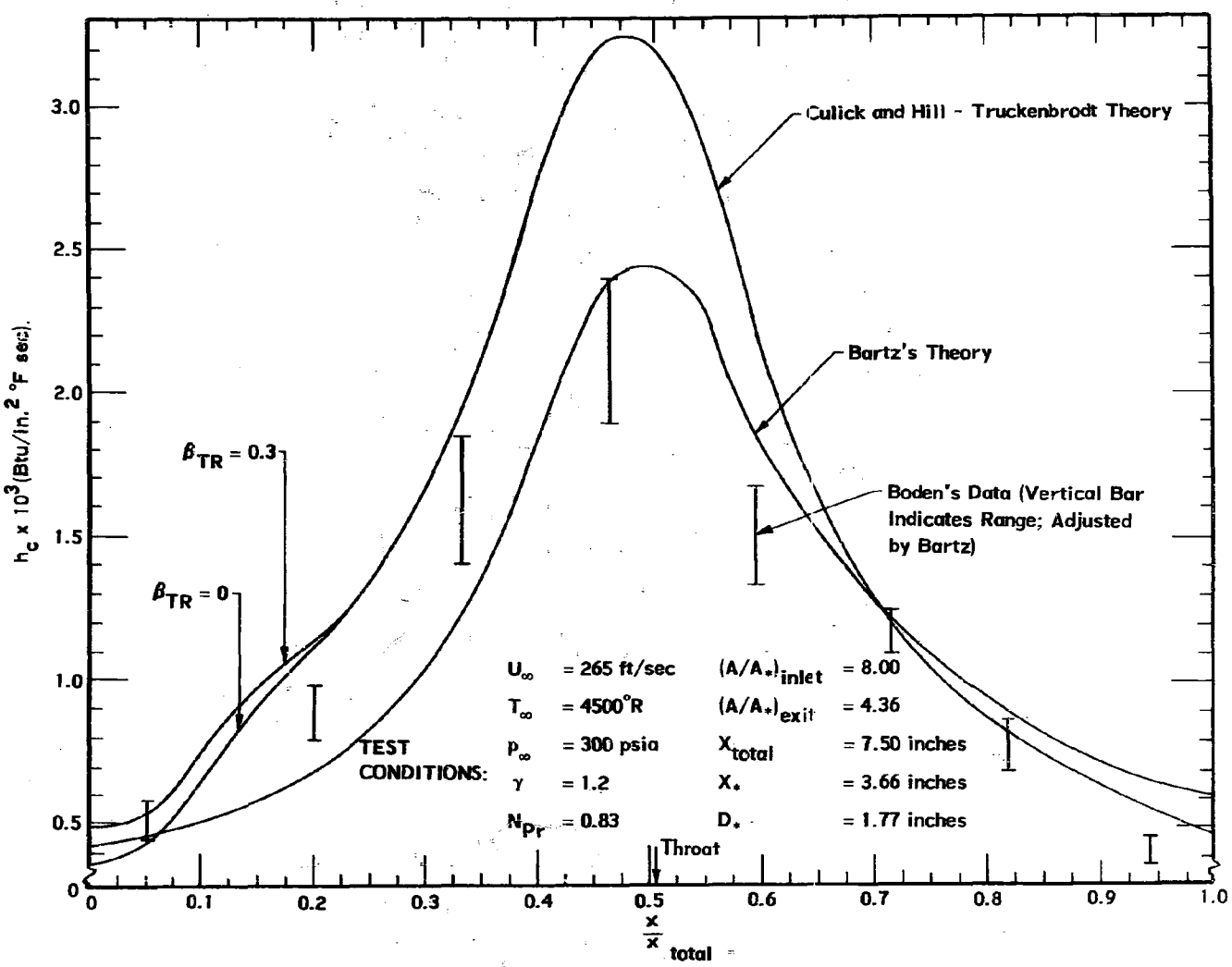


Fig. 5. The Forced Convective Heat Transfer Coefficient: Culick and Hill - Truckenbrodt theory, Bartz Theory and Experimental Results

The calculation is started by assuming (or calculating) the momentum thickness at the point of transition from laminar to turbulent flow, θ_{TR}, and the Falkner-Skan pressure gradient parameter at transition, β_{TR}. For most practical cases, θ_{TR} and β_{TR} can be assumed to be zero. Knowing the propellant gas velocity at the transition point, U_{TR},* and the kinematic viscosity of the propellant gas as a function of temperature and pressure, the momentum thickness Reynolds number at the transition point, $N_{Re_{\theta_{TR}}}$ may be calculated from

$$N_{Re_{\theta_{TR}}} \equiv \frac{U_{TR}\,\theta_{TR}}{\nu_{\infty}} \tag{14}$$

Tables of β_i versus $H_i\Big|_{Laminar}$, $N_{Re_{\theta_{TR}}}$ versus ΔH_i,

where $$\Delta H_i \equiv H_i\Big|_{Laminar} - H_i\Big|_{Turbulent} \tag{15}$$

and $H_i\Big|_{Turbulent}$ versus $L_i\Big|_{Turbulent}$

are known and constitute part of the general input to the program. The absolute viscosity is estimated from Sutherland's law, which is

$$\mu = \mu_* \left(\frac{T}{T_*}\right)^{3/2} \left(\frac{T_* + 198.6}{T + 198.6}\right) \tag{16}$$

Sutherland's constant for air has been assumed to be applicable. The propellant gas density may be assumed, in a first approximation, to conform to the perfect gas law, thus,

$$\rho = \frac{PM_g}{R_o T} \tag{17}$$

The velocities, pressures, and temperatures at the edge of the compressible boundary layer are either measured in model experiments or computed as previously suggested. These values, along with the Prandtl number and the isentropic expansion coefficient of the propellant gas comprise the input quantities required to initiate the stepwise calculation from the transition point.

---

*The transition point is located near the stagnation point dividing Zones I and II, and thus the grain port velocity provides an estimate of U_{TR}.

The transformations of Culick and Hill, which are as follows:

$$U_i = \sqrt{\frac{T_\infty}{T}}\, U \tag{18}$$

$$\theta_i = \left(\frac{T}{T_\infty}\right)^3 \theta \tag{19}$$

$$x_i = \int \left(\frac{T}{T_\infty}\right)^{[(\gamma + 1)/2(\gamma - 1)] + 5/6} dx \tag{20}$$

and

$$C_f = C_{f_i} \left(\frac{T}{T_\infty}\right)^{5/6} \tag{21}$$

are applied to the input information to allow their subsequent use in the appropriate incompressible equations (see reference 17, equations 22.39, 22.32, and 22.31, as well as equation 42 of reference 16). The compressible skin friction coefficient, C_f, as a function of axial position and propellant burn-back time is thus obtained. Using the modified Reynolds analogy (21), (22),

$$h_c = \frac{\rho U C_p}{(N^*_{Pr})^{2/3}} \left(\frac{C_f}{2}\right) \tag{22}$$

where N^*_{Pr} is evaluated at the Eckert reference temperature, (12) defined by

$$T^* = \frac{T_w + T}{2} + (0.22)\, r' \frac{(\gamma - 1)}{2} M^2 T \tag{22a}$$

(r' is the recovery factor (see Eq. 25)),

and C_p is the isobaric specific heat, assumed to be a constant for the propellant gas system of interest and calculated from

$$C_p = \frac{\gamma R_o}{(\gamma - 1) Mg} \tag{23}$$

Comparison With the Bartz Equation. The solution obtained from Culick and Hill's modification of the Truckenbrodt analysis was programmed. Using this computer program, a sample calculation was made for comparison with the rocket motor heat-transfer data reported in references (10) and (11). It should be observed, at least for the data of these references that (a) Bartz's method tends to slightly over-predict the heat transfer in the exit portion of the nozzle, (b) the experimental results showed a maximum deviation of 80 percent above and 45 percent below the analytical values at the throat, and (c) the experimental results were found to be as much as 100 percent

above the analytical values in the entrance to the nozzle (note Zietland's prediction (23) of the importance of entrance effects in this region).

Figure 5 presents experimental results for a test performed using the same rocket motor as was used by Welsh and Witte. These data were used as a comparison between results from the Culick and Hill - Truckenbrodt analysis and the Bartz analysis (24). Both sets of analytical results, together with the experimental results, are shown. It can be seen that the heat-transfer coefficient calculated by the present analysis agrees well with the experimental data in the exit portion where the coefficient based on Bartz's method is slightly high. At the throat, the present method gives results about 30 percent above both Bartz's results and the experimental data, and it predicts values higher than the Bartz equation in most of the entrance region, although farther upstream there is a crossover in the results (for $\beta_{TR} \equiv 0$). The set of experimental results in Figure 5 lie between the two analytical predictions. The data of Welsh and Witte for the entrance and throat region are often above the Bartz equation result and tend to be nearer the value predicted by the present method. Although there is considerable scatter in the experimental data, it can be seen that the present analysis compares well in all sections of the aft closure, throat and exit cone.

The importance of the assumed or calculated transition quantity, β_{TR}, on the computed forced convective heat transfer coefficient, h_c, is also shown in Figure 5. For a large variation in values of β_{TR}, the importance of this parameter rapidly diminishes with distance through the aft-closure, and has a negligible influence on h_c beyond about one-half of the way through the aft-closure entrance section of the nozzle system for which the curves and data in Figure 5 were obtained. Similar results were obtained when the θ_{TR} was varied by an order of magnitude (0.001 ft to 0.0001 ft).

Using the coefficient determined by the method of Culick and Hill, the convective heat flux can be obtained by the equation

$$q_c = h_c[T_{aw} - T_w] \tag{24}$$

providing a reasonable estimate of T_w can be made. The adiabatic wall temperature (or recovery temperature) can be calculated from

$$r' = \frac{T_{aw} - T}{T_o - T} \tag{25}$$

where $r' = N_{Pr}^{1/3}$ for turbulent flow over flat plates (25), (21). For rocket effluent gases r' is approximately 0.9 (i.e., $N_{Pr} \approx 0.7$), and therefore, since the Mach number is low (i.e., $T_o \approx T$) in the aft-closure region,

$$T_{aw} \approx T_o \tag{26}$$

or

$$q_c = h_c [T_o - T_w] \tag{27}$$

Particle Effects

The existence of alumina particles in the exhaust gases of a solid propellant rocket create a severe environment for the materials which are used to inclose the aft end of the rocket. Ungar (26) has demonstrated that if such gas-particle flows impinge on ablative surfaces, the rate of material removal is increased over that for a similar gas flow. The primary reasons for this severe condition leading to the increased material loss are the larger heat transfer rates caused by the paricles and the mechanical erosion of the surface due to particle-surface interaction. The increase in heat transfer is a threefold effect: first, the particles interact with the boundary layer, thus increasing the convective heat transfer; second, the particles penetrate through the boundary layer and transfer a portion of their kinetic and thermal energy to the wall upon collision; and third, the presence of a particle cloud effect increases the radiant heat flux to the surface. The significance of the mechanical erosion is believed to be the removal of the thermal insulating char layer which most materials form and the subsequent removal of virgin material through mechanical erosion.

The first two heat transfer effects are very difficult to evaluate on the basis of current knowledge. As a result their total effect will be lumped together and consideration will be given to the additive result. The method proposed is to use proportionality factors analogous to the accommodation factors defined in free molecular flow in order to prescribe that amount of energy which the particles contain that is transferred to the wall during collisions. Thus one can write the heat flux due to particle impingement as

$$q_p = \dot{m}_p \left[\alpha_t c(T_p - T_w) + \alpha_k \frac{U^2}{2g_c J}\right] \tag{28}$$

where α_t and α_k are these proportionality factors. Since the particle velocities are relatively small in the aft-closure and the particle enthalpies are large, the total kinetic energy which the particles have is small in comparison with the thermal energy. Thus one can simplify by allowing α_k to equal 0 and considering only the thermal accommodation coefficient. It has been shown by Farbar and Morley (27) that the heat transfer rate is not increased appreciably at particle to gas loadings of ϕ less than 1.0 for straight pipe flow. Since the particle loadings in most rocket motors are considerably less than $\phi = 1.0$, it can be assumed that the heat transfer due to particle impingements will only be significant in those regions where the flow must go through large turning angles as discussed previously. Consequently simulated flows with particles are necessary to delineate these areas. Ungar (26) has analytically examined results obtained using materials that form liquid layers and successfully correlated his results. However, these results are not general enough to extend to other systems and will not be discussed further. Since very little information is available as to the values of the thermal and kinetic accommodation coefficients, the true importance of this heat flux cannot be evaluated. However, it is apparent that it may be sizable at positions of large turns and may be the factor which accounts for the high erosion rates in small sections of the aft-closure.

The presence of a large number of alumina particles in the solid rocket exhaust gases increases the importance of heat transfer by radiation. Investigators of heat transfer in liquid propellant rocket nozzles have assumed that the heat transfer by radiation from the exhaust gases was negligible. This assumption is based on the fact that the hot gases do not radiate an energy continuum but radiate only in small finite frequency bands. Although in these bands the radiation may be of black body intensity, the total radiancy from the gas is much less than the black body value. The presence of a solid phase which is emitting a continuum increases the overall radiation level. If this cloud is sufficiently dense, the effective emissivity can approach unity. Recently some experimental work has been carried on by Carlson (28) to measure the radiance of particle clouds in the exhaust of solid rockets. Further, a method (29) has been reported to quantitatively calculate on the basis of theory the emissivity of particle clouds. Both the experimental results and the theoretical method will be discussed with particular emphasis on their applicability to alumina clouds as found in rocket exhausts.

The experimental results reported by Carlson (28) are for the radiation from MgO clouds in an RP-1/gaseous oxygen flame. A slurry of RP-1 and MgO particles was combusted with gaseous oxygen in a rocket motor and the effluent expanded through a nozzle. The radiation of the cloud was measured at an expansion ratio of 5.0. The throat diameter was 1.25 inches. In the tests both the fuel-to-oxidizer ratio (the flame temperature) and the weight of particles-to-weight of gaseous effluent were varied. The results are shown in Figure 6. Since the optical properties are relatively insensitive to particle temperatures, the absorptivity is not expected to be highly dependent upon the oxygen-to-fuel ratio, and the data points of Figure 6, therefore, include several values of the oxygen-to-fuel parameter. The size distribution of the MgO particles is given in Figure 7.

The results shown in Figure 6 are correlated by the expression

$$\epsilon = 1 - e^{-k\phi} \tag{29}$$

the form of which can be arrived at from theory. It has been shown (30) that the absorptivity of a particle cloud is given by the following relationship

$$\alpha = 1 - \exp\left[X \int_o^\infty \pi R_p^2 N(R_p)\, Q_{ext}(R_p)\, dR_p \right] \tag{30}$$

where

$$Q_{ext}\ (R_p) = \frac{\sigma_{abs} + \sigma_{sc}}{\pi R_p^{\,2}} = \frac{\sigma_{eff}}{\pi R_p^{\,2}} \tag{31}$$

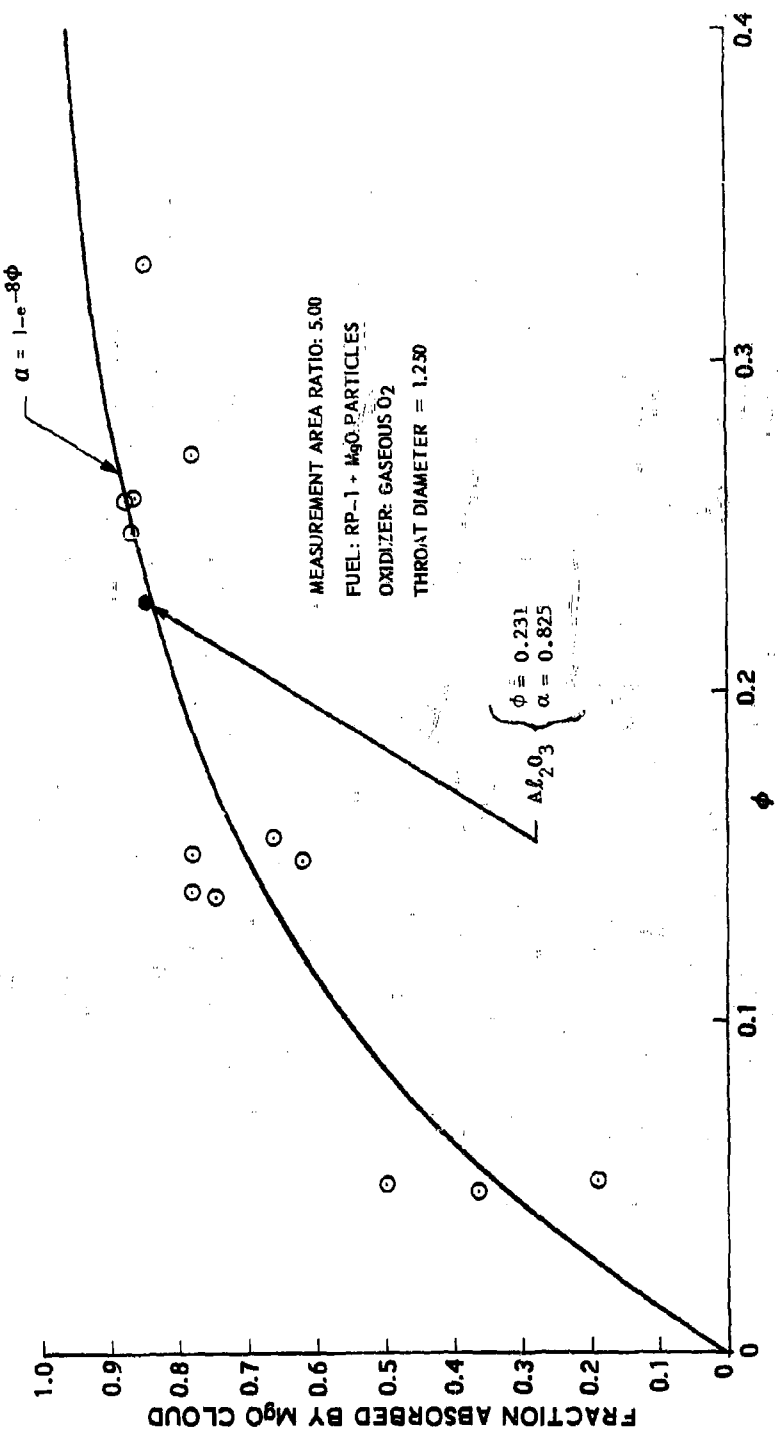


Fig. 6. Absorptivity of Rocket Exhaust Due to Presence of MgO Cloud - Wavelength of Measurement = 5911 Angstroms

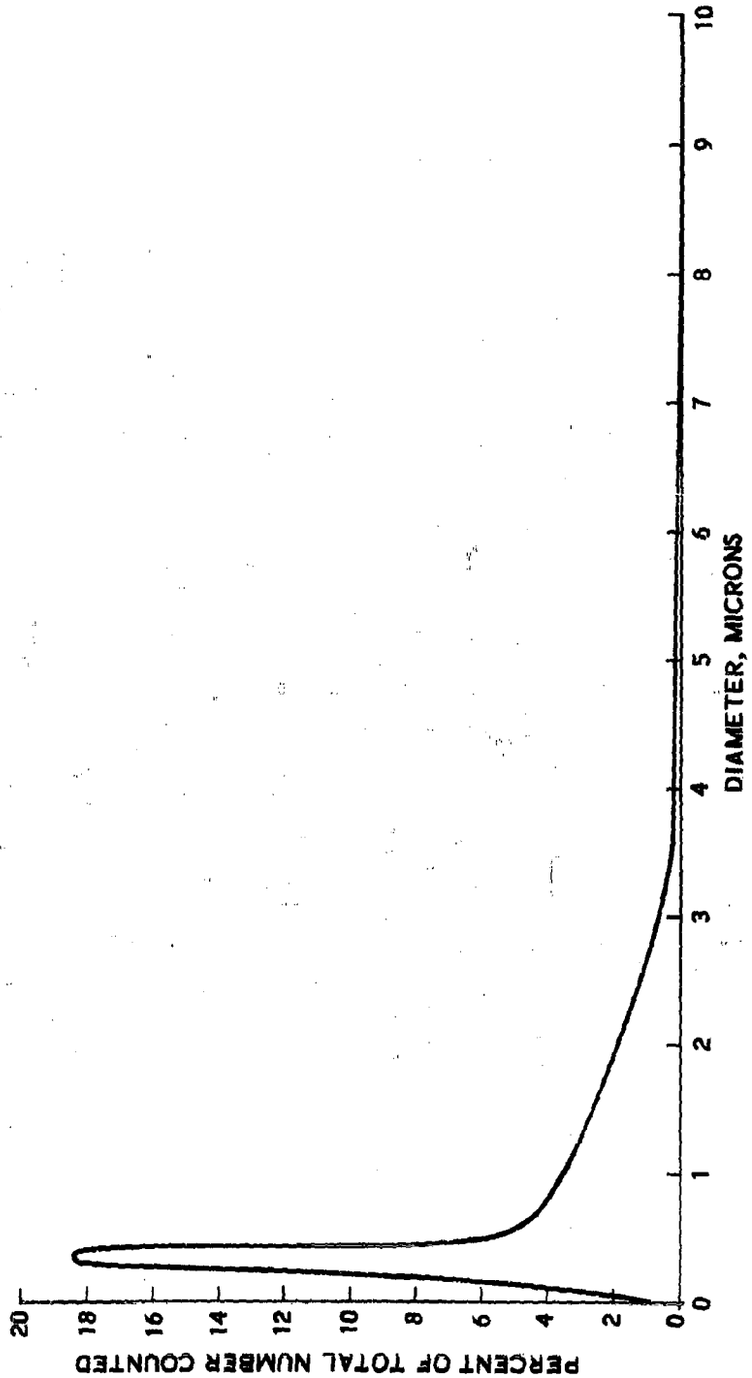


Fig. 7. Distribution of MgO Particle Sizes Used in Velocity-Lag Tests

Assuming that all the particles in the cloud are of the same radius this expression can be integrated to yield

$$\alpha = 1 - \exp(NX\sigma_{eff}) \tag{32}$$

Also

$$\phi = \frac{N(\frac{4}{3}\pi R_p^3)\,\rho_{particle}}{\rho_{gas}} \tag{33}$$

where ϕ is the ratio of the mass flow rates of particles to gas. Substituting for N in Eq. (32) gives

$$\alpha = 1 - \exp\left[\left(\frac{6}{\pi D^3 \rho_s}\right)\rho_g\, X\phi\sigma_{eff}\right] \tag{34}$$

If exhaust clouds contain the same chemical species of particles (i.e., same optical properties), having the same size distribution, and if radiancy measurements are made at the same area ratio, Eq. (34) can be reduced to the expression used to correlate the results of Figure 6.

If these same conditions apply to two different motors, the following ratio is obtained from Eq. (34):

$$\frac{\ln(1-\alpha_1)}{\ln(1-\alpha_2)} = \left(\frac{\rho_{g_1}}{\rho_{g_2}}\right)\left(\frac{X_1}{X_2}\right)\left(\frac{\phi_1}{\phi_2}\right) = \eta \tag{35}$$

or

$$\alpha_1 = 1 - (1-\alpha_2)^{\eta} \tag{36}$$

Eq. (36) allows the estimation of the emissivity of a given cloud based on an experimentally determined emissivity of another cloud. In order to extend the applicability of the MgO data, Carlson has also fired the slurry motor using alumina rather than MgO particles. The results of these tests have not been reported to date; however, one point which has been measured is included in Figure 6. Although the amount of data is quite small, the good agreement between the alumina emission and MgO emission lends some support to the use of this data to approximate the emissivity. By using the results of this one test and Eq. (36), the following relationship is obtained:

$$\alpha = 1 - 0.175^{\eta} \tag{37}$$

If one now can consider a motor having a 20% aluminum loading; that is, a ϕ of 0.67, it is seen from Eq. (37) that an emissivity of 0.99 will be obtained for a diameter of about 2.7 inches. Although the results of this calculation

do not allow one to accurately determine the emissivity of a particle cloud at any position for any motor, it is felt that it does show that the effective emissivity of the particle clouds in motors of the size of current interest is close to unity.

A method of calculating the radiancy of particle clouds has been carried out by Stull and Plass (29). The particular problem solved by them was for solid carbon particles in flames; however, the method is applicable to the solid propellant case if one knows the optical properties of the alumina at the elevated temperatures encountered. The method that is outlined by Stull and Plass allows the determination of the scattering and absorption cross-sections for the individual particles as calculated from the Mie theory. A dispersion equation is derived which represents the optical properties of the carbon at the elevated temperatures. Based on these values an expression is obtained for the radiation intensity emitted by a large number of dispersed particles which includes all higher order scattering processes. Stull and Plass demonstrate this technique for particle clouds containing single sized particles (Distribution I) and for two different size distributions as shown in Figure 8 (Distributions II and III). Figure 9 shows typical results obtained in this study where ℓ is the optical cross section of the cloud. As one might expect, at higher optical cross sections, the radiation intensity or emissivity approaches black body conditions. It should also be pointed out that this method is limited to particle distributions in which all the particles are at the same temperature. It has been shown that, as the particles are accelerated in the nozzle, a velocity and temperature lag develops which is a function of the diameter of the particles (31). As a result not only is there a particle size distribution in the nozzle but there also is a particle temperature distribution which is a function of the size distribution. Before the method of Stull and Plass can be applied throughout the nozzle, the derivation must be revised to include this temperature variation.

The principle reason that this theoretical approach is not utilized at present is the lack of data as to the optical properties of alumina at temperatures that are typical of those found in rocket exhausts. There are several programs under consideration that will lead to a better understanding of the overall problem including the evaluation of the optical properties of alumina in exhausts. However, until this work is completed only empirical predictions of the cloud emissivity based on an insufficient quantity of data are available. Another problem in the use of the empirical data is that the maximum intensity for the radiation occurs at smaller wave lengths than the corresponding black body intensity maximum (see reference 29). As a result, the concept of an effective emissivity is not valid; that is, the cloud does not act as a "gray" body. Although this is not a serious error for clouds whose emissivity is near unity, it could introduce error at lower emissivities. Also Stull and Plass show that the emissivity of a cloud of size distributed particles is not the same as that for a cloud of the same number of particles, all having the diameter that is the mean of the distributed cloud. Carlson (28) pointed out that the radiation from a cloud having a temperature distribution as discussed previously radiates like a single sized cloud of particles whose diameter and corresponding temperature was greater than the distributed cloud mean. In short, the radiation from clouds of particles having a size

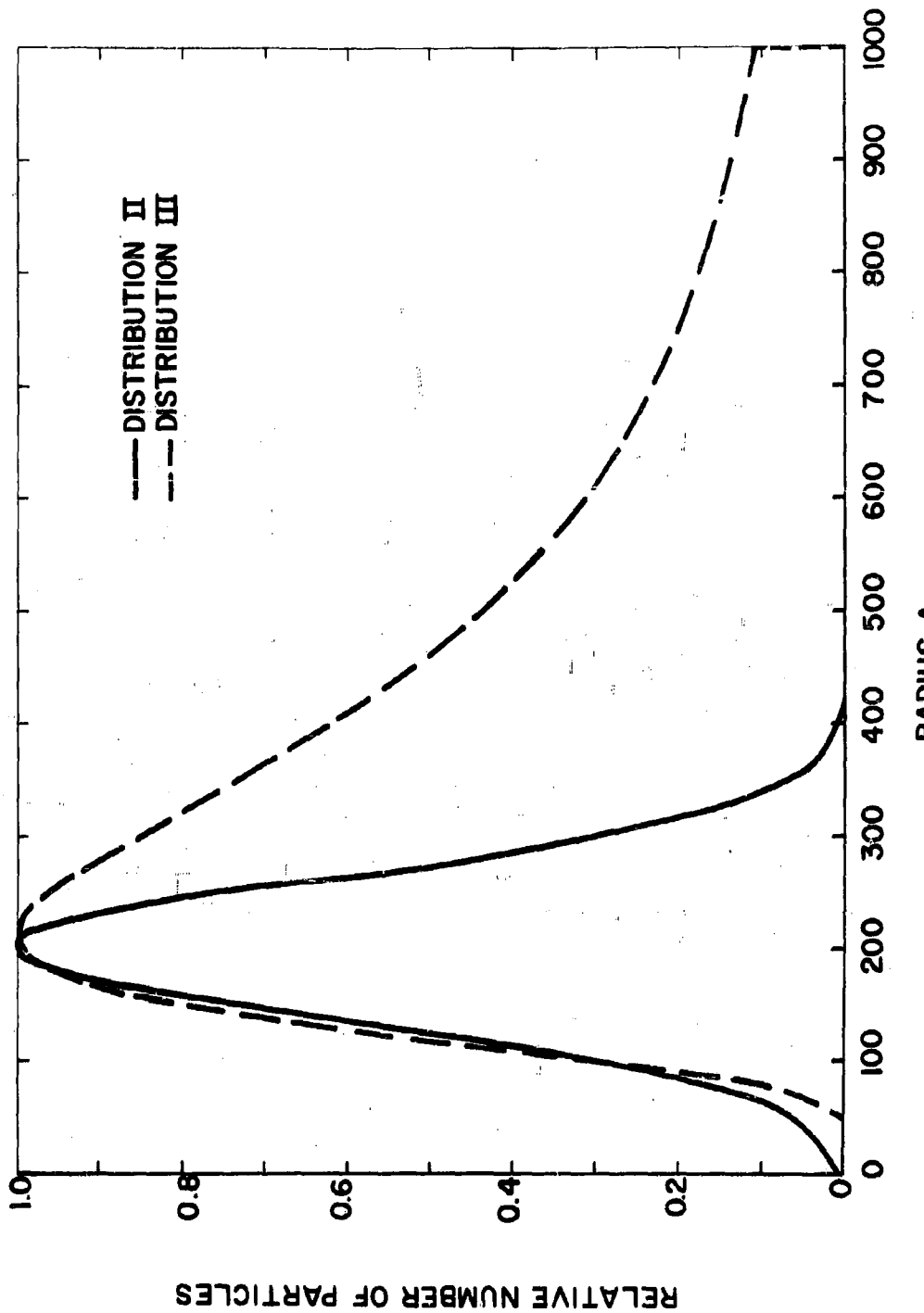


Fig. 8. Two Particle Size Distributions Used for the Calculation of the Emissivity of Carbon Particles

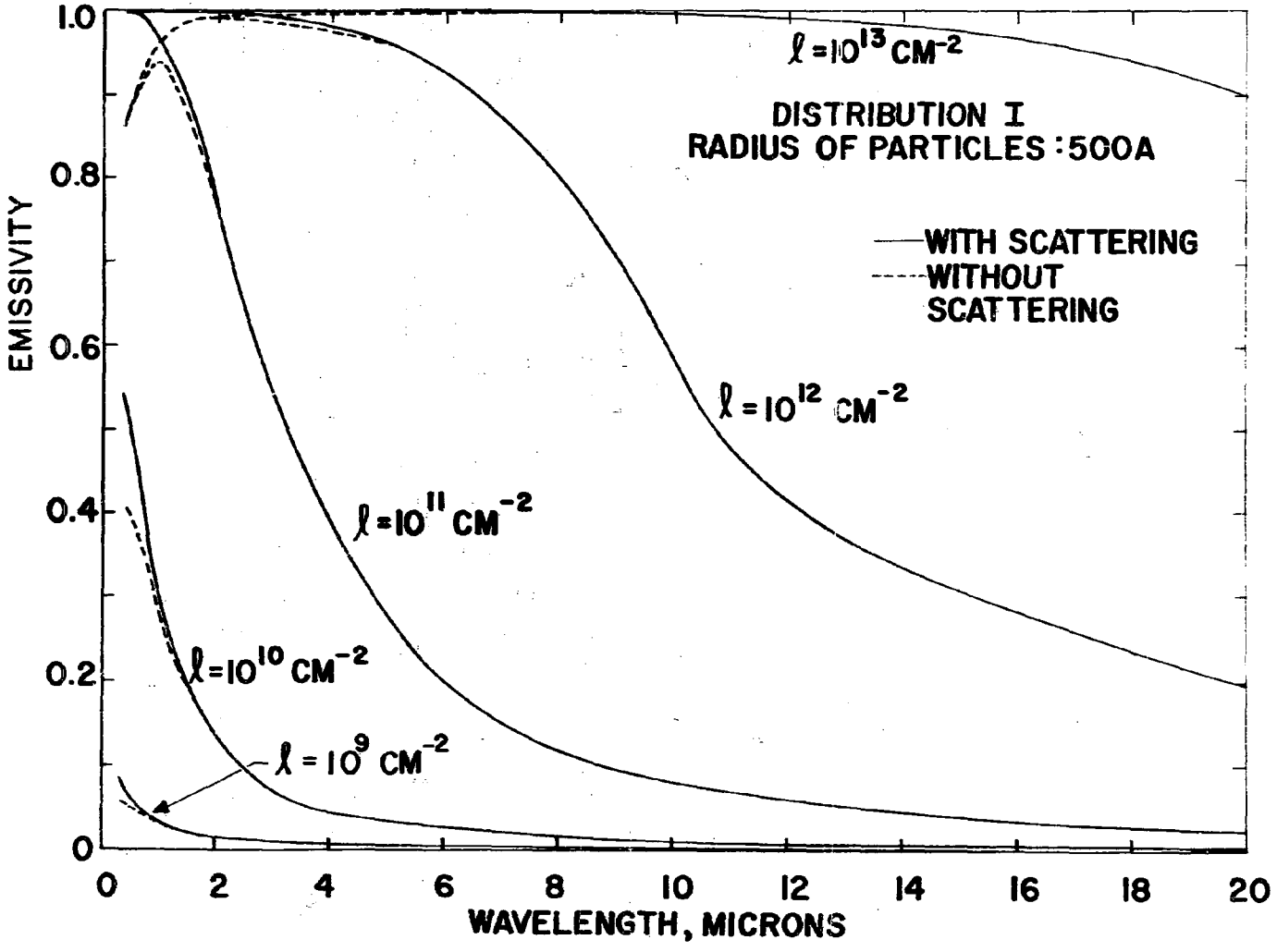


Fig. 9. Emissivity as a Function of Wavelength for Carbon Particles of 500 A Radius

distribution does not bear a simple relation to the radiation from a similar single sized cloud of particles of the distribution mean diameter.

In the aft-closure region of most current motors many of these limitations have only a minimal effect on the radiative heat flux, but they are reported as a warning against an indiscriminate application elsewhere. Considering only the aft-closure region of highly loaded metallized solid propellant motors of large diameters (typical diameters are greater than six inches), it has already been shown that the cloud emissivity is near unity and the particles are all in equilibrium with the gas at the gas stagnation temperature. Consequently, the radiant flux can be written as

$$q_r = \epsilon\sigma\left[T_o^4 - T_w^4\right] \tag{38}$$

where

$$\epsilon = \left(\epsilon_{wall}^{-1} + \epsilon_{cloud}^{-1} - 1\right)^{-1} \tag{39}$$

The most important effect of particle impingement on the wall is that of mechanical erosion. As discussed previously this mechanical erosion may well remove the thermal insulating char layer that most aft-closure insulation materials form during firing. Once this insulating layer has been removed, the heat transfer to the virgin material is increased, thus increasing the ablation rates. Also, the particles can then mechanically erode the virgin material. The significance of this model is that it is not necessary to consider the ablation in evaluating the mechanical erosion, so that cold flow tests with particles impinging upon cold surfaces will give an estimate of the erosiveness of the stream. By relating the amount of particle impingement at a particular point to these cold flow tests, one can acquire a reasonable estimate of the virgin material removal due to erosion. Moreover, a simplified model based on no char layer ablation can be analyzed in order to estimate the amount of virgin material removal due to thermal degradation of the insulating materials. Again the validity of such a model is in question, but by pointing out the possibility of the existence of such a model it is hoped that emphasis will be placed in future experimental work upon the evaluation of these particle effects.

Ablation Rate Determination

In Zone II the insulation material ablation rates are large and will be strongly influenced by particle effects. The variation with time of the hydraulic radius of the grain port and the grain port design itself are important factors which cause the fluid flow in Zone II to be three dimensional and unsteady. It is apparent that when this complexity is coupled with that of two-phase flow and ablation, no analytical methods are available for the prediction of the heat and mass transfer rates at the wall. The approach used here is to employ an effective heat of ablation, Q*, which is in part a material property, in order to estimate the insulation material removal rate. The Q* concept has been extensively discussed in the literature of ablation (32), (33), (34), (35), (36), (37), (4). The effective heat of ablation is defined by (35),

$$Q^* = \frac{\text{heat transfer to a non-ablating calorimeter}}{\text{total mass ablated}} \tag{40}$$

where the calorimeter is assumed to have the same surface temperature, emissivity, and catalytic efficiency as the ablating material. If, in a laboratory determination of Q* for an ablating material which forms a char layer (which may or may not be structurally weak), care is taken to retain all the solid material (virgin elastomer plus char particles and layer) by containing the sample in a short length of a cylinder , say, then the difference between the before and after weight measurements of the ablating sample will give a value for the total mass of gas evolved ($\dot{m}_g$) and lost as a result of the test. From a calibration measurement using a non-ablating calorimeter (with provision for the measurement of the surface temperature by optical pyrometry, say,) the total heat flux to the sample may be determined for known arc-jet or torch conditions. Thus, on a unit time basis, the Q* so measured can be related to the energy fluxes existing in Zone II of the aft-closure by,

$$Q^* = \frac{q_c + q_r + q_p}{\dot{m}_g} \tag{41}$$

where, for char forming elastomeric insulating materials, the blowing effect (which decreases the net heat transfer to the wall) has been considered small relative to the sum of the fluxes given in the numerator. Blowing effects will probably be minimized, considering the secondary flows and particle effects, which would tend to counteract any boundary layer thickening caused by mass transfer. Nothing is now known concerning the gas composition given off at elevated temperatures by typical elastomeric insulation materials, though techniques for obtaining this information have been developed (38) for measuring the gas composition given off by plastic ablators and should be equally applicable to elastomeric materials. It is more appropriate to evaluate the importance of this effect when this information becomes available. The linear regression rate, r, is related to $\dot{m}_g$ as follows:

$$r = \frac{\dot{m}_{total}}{\rho_{in}} = \frac{\dot{m}_g + \dot{m}_c}{\rho_{in}} \tag{42}$$

where $\dot{m}_c$ is the mass of porous char formed per unit surface area and per unit time. Since the virgin material is homogeneous, then that portion of it which is transformed to the gas phase is proportional throughout the insulation material to that which forms the porous char. Also, since the char layer is assumed to be swept away as soon as it is formed, the following proportion,

$$\dot{m}_c = \frac{\rho_c}{\rho_{in}} (\dot{m}_g + \dot{m}_c) \tag{43}$$

may be used to eliminate $\dot{m}_c$ from Eq. (42) and thus obtain,

$$r = \frac{\dot{m}_g}{\rho_{in} - \rho_c} \tag{44}$$

where ρ_c is the density of the porous char ($\sim \rho_1$ in Grosh nomenclature). Eqs. (27), (28), (38), (41), and (44) can be combined to yield

$$r = \frac{[\sigma\epsilon(T_o^4 - T_w^4) + (h_c + \alpha_t \dot{m}_p\, c)(T_o - T_w)]}{Q^*(\rho_{in} - \rho_c)} \tag{45}$$

Using an estimate or measurement of T_w (see reference 39), and knowing Q^* as a function of $q_{total} = q_c + q_r + q_p$) for the particular insulation material of interest, then r may be calculated for each wall position in Zone II. Of course, h_c will depend on the static pressure gradient or U's employed as input to the Culick and Hill analysis. Methods such as cold or hot flow modeling need to be developed which will enable the prediction of reasonable estimates of this information as well as the (α_t $\dot{m}_p$) at each position in Zone II as a function of time.

INPUT INFORMATION

In order to successfully apply the preceding considerations to the determination of the heat transfer rates and the insulation material removal rates in an aft closure, the following conditions and properties must be known (or estimated):

a. General flow, gas and particle properties:

(1) Rocket aft closure, and grain port nozzle geometry, and the change of grain port geometry with time.

(2) Grain port velocity (or C*).

(3) Propellant gas properties (thermodynamic and transport) including chemical state and initial particle mass fraction (neglecting the influence of the particles on the gas transport properties).

(4) Chamber pressure as a function of time including the total firing time, t_e.

(5) Chamber temperature (flame or stagnation temperature).

(6) Fluid velocities at edge of boundary layer (note that for $N_{Pr} \sim 1$ fluids, the momentum and thermal boundary layer thicknesses may be taken to be identical) at all times and positions.

(7) Inside wall (gas-char or gas-reaction zone interface) temperature as a function of axial position (note that in this analysis, a constant wall temperature has been assumed for each zone).

(8) The optical properties (i.e., complex refractive index) of the metal (oxide) particles at the chamber and aft-closure temperatures of interest as well as the number density of particles and particle size distribution appropriate for the given flow situation.

(9) Particle impingement rates ($\dot{m}_p$) at various locations and times in the aft-closure for given conditions 1 through 7.

(10) The energy accommodation (α_t) of the particles upon collision with the insulation wall for given conditions 1 through 7.

(11) A measure of the mechanical erosiveness of particles on both virgin insulation materials and their char layers.

b. Insulation material properties:

(1) The thermal conductivity, mass density and specific heat for the virgin material at a temperature between the ambient and mean reaction zone temperature.

(2) The thermal conductivity, mass density and specific heat of the porous char at a temperature between the mean reaction zone temperature and T_w (or the porosity of the char and the corresponding properties of the solid matrix material which makes up the char).

(3) The overall heat of pyrolysis and phase change (L_1) and the mean reaction zone temperature (T_1) and their possible dependence on the surface temperature.

(4) The composition, k, C_p and mass density of the gas mixture which transpires through the porous char layer at a temperature between the mean reaction zone temperature and the surface temperature (T_w) at the chamber pressure.

(5) Q* values for the ablative insulation material measured under specially designated experimental conditions (as a function of total incident heat flux and chamber pressure)

In regards to the present status of knowledge concerning each of the needed conditions and properties listed under a and b, the following information is not currently available and must at best, be crudely estimated (in some cases by techniques suggested in the previous sections): a-6 through a-11 with the possible exception of a-7 (39) and a-8 (28); and for elastomeric insulation materials, b-2 through b-5. A resonable analytical a priori determination of heat transfer in an aft closure requires knowledge of these properties.

OVERALL CALCULATION PROCEDURE

The analyses indicated above are integrated into a logical sequence of calculations to give the final insulation profile. The thickness of insulation that is either charred or removed in the recirculation zone, Δ_I, can be calculated as a function of time. Either by modeling the flow or by approximations based on theory (1), the location of the stagnation point can be prescribed as a function of time. Thus the length of time during which a particular point on the wall is in the secondary flow zone is known, and also the amount of material charred or removed during that time. After the stagnation point passes a point, it is assumed that the char is essentially removed and then the "fast" ablation continues. Using the Q* concept, one can write

$$r(t) = \frac{q_p(t) + q_c(t) + q_r}{Q^*(\rho_{in} - \rho_c)} \tag{46}$$

If the values of $q_p(t)$ and $q_c(t)$ can be determined, $r(t)$ can be determined and during fast ablation, the material removed is then

$$\Delta_{II} = \int_{t_s}^{t_e} r(t)\, dt \tag{47}$$

To complete the determination of the amount of material removed, experimental data is needed as to the amount of mechanical erosion to be expected. Finally allowance must be made for the thermal penetration into the insulation at the end of the firing. This can be approximated from a steady state solution of the semi-infinite, one-dimensional conduction equation assuming a constant surface regression rate and a maximum allowable back wall temperature, T_{max}. The result of this is (see Baer and Ambrosio (40) Eqs. (30) and (31))

$$\Delta_p = -\frac{\alpha}{r} \ln\left[\frac{T_{max} - T_i}{T_w - T_i}\right] \tag{48}$$

where r is the regression rate at burnout. The final insulation thickness required at a point is then

$$\Delta = \Delta_I + \Delta_{II} + \Delta_e + \Delta_p \tag{49}$$

where Δ_e denotes the total material removed due to mechanical erosion. For convenience of ordering the calculations the insulation wall positions at which material removal and thermal penetration are calculated are located along a line on the wall formed by the intersection of a plane through the axis and the wall. Circumferential variations can be included by rotating this plane and completing another set of calculations. The complete calculation outlined herein is dependent on the ability to describe the various phenomena discussed as functions of time and position. Figure 10 shows in schematic form the calculation procedure.

CONCLUSIONS

Although the prediction of insulation requirements in the aft-closure of a metallized solid propellant motor is of considerable importance, it is impossible to develop a usable, inclusive analysis of the problem because of the lack of information concerning many of the phenomena. In this report a calculation technique is developed which is a simplification of the total problem. For instance, effects such as the influence of boundary layer reactions upon heat transfer, the reduction of the convective flux due to blowing, and wall roughness have not been considered. However, an attempt has been made to include those effects that are felt to be of prime significance. A method of handling each of these effects is suggested and then the missing, necessary information is identified. In conclusion, the method described herein should provide relatively accurate predictions of aft-closure insulation requirements when the following data becomes available:

(1) Pressure and velocity at the edge of the boundary layer at any time and position in the aft closure.

(2) Particle impingement rates at various locations and times in the aft closure.

(3) Methods for estimating the accommodation of the energy of the particles upon collision.

(4) Estimation of the mechanical erosiveness of particles on both virgin insulation materials and their char layers.

(5) The complex refractive index of the metal oxides at elevated temperatures.

(6) The physical properties of the virgin ablators and their char layers.

(7) Q* values for ablators measured under specially designated experimental conditions.

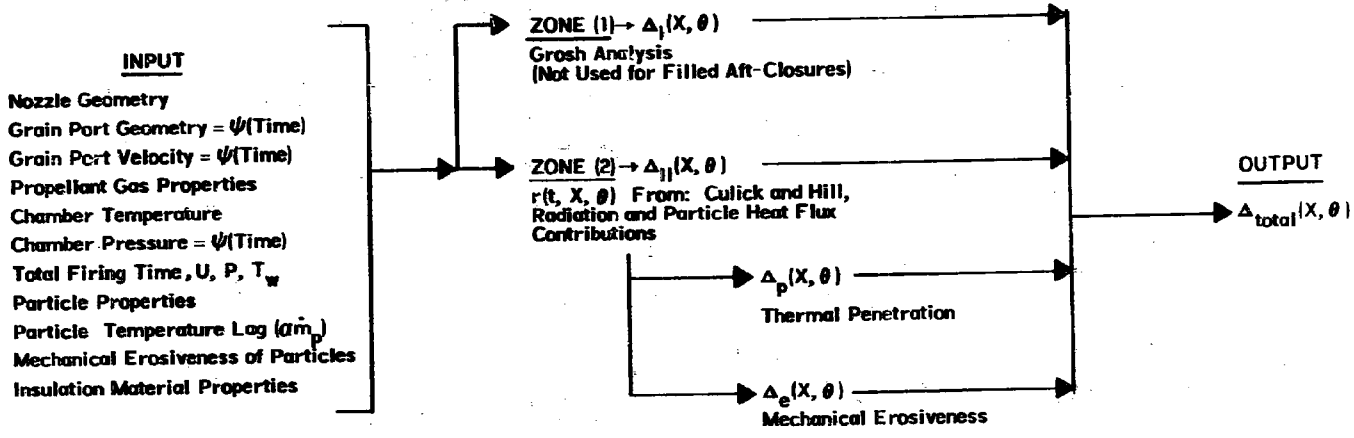


Fig. 10. Schematic Calculation Procedure for Determining Insulation Thicknesses

NOMENCLATURE

A	local cross-sectional area of flow
A_*	cross-sectional flow area at nozzle throat
C_f	local skin-friction coefficient, $C_f \equiv \frac{2\tau_w}{\rho U^2}$
C_p	specific heat of the gas at constant pressure
C^*	characteristic velocity (see reference 11)
c	heat capacity of particles
D_*	diameter at the nozzle throat
g	gravitational acceleration (32.17 ft/sec^2)
g_c	gravitational conversion factor (32.17 $\frac{lb}{lb_f}\ \frac{ft}{sec^2}$)
H	boundary-layer shape parameter (see Reference 17)
h_c	local convective heat-transfer coefficient
J	mechanical equivalent of heat (778.2 $\frac{ft\ lb_f}{Btu}$)
L	Truckenbrodt's shape factor (see Reference 17)
ℓ	optical cross-section of cloud
M	Mach number at the edge of the boundary layer
M_g	molecular weight of propellant gas mixture flowing through aft closure
$\dot{m}_g$	rate of gas evolution of ablating insulation material per unit surface area
$\dot{m}_p$	rate of particle impingement per unit area
k	proportionality constant
N	particle number density
N_{Pr}	Prandtl number of the propellant gas mixture
P_c	chamber pressure
Q_{ext}	extinction coefficient

Q^* effective heat of ablation for insulating material (see Eq.41)

q_c convective heat flux per unit area

q_r radiative heat flux per unit area

q_p energy exchange per unit area as a result of particle impingements

r linear regression rate of insulating materials

r' recovery factor, taken to be $N_{Pr}^{1/3}$ for turbulent flow

r_c radius of wall curvature at the nozzle throat

R local radial distance from the rocket motor centerline

R_p radius of particles

R_o Gas Law constant = 1/987 Btu/lb mole °R

T static gas temperature at the edge of the boundary layer

T_i initial temperature of the insulation

T_o stagnation temperature of gas free stream

T_p temperature of particles

T_{aw} adiabatic wall temperature

T_w temperature of wall

t_e total firing time

t_s time at which flow stagnation point moves past a given station after ignition

U local edge-of-boundary-layer velocity

v_{p_i} volume of typical particle

X cloud path length or axial position in nozzle measured from the start of the aft closure

x distance along the surface in the direction of flow

Greek Symbols

α_t impinged particle's thermal-accommodation coefficient

β Falkner-Skan pressure gradient parameter (see reference 17, p. 132)

γ ratio of specific heat at constant pressure to that at constant volume (frozen)

ϕ ratio of the mass flow rate of particles to the mass flow rate of the gas

ϵ effective emissivity (see Eq. 39)

σ Stefan-Boltzmann constant, Eq. (38)

$\sigma \quad \left(\frac{\rho_{am}}{\rho}\right)^{0.8} \left(\frac{\mu_{am}}{\mu_o}\right)^{0.2}$ in Eq. (12)

σ_{abs} absorption cross-section

σ_{eff} effective cross-section

σ_{sc} scattering cross-section

ρ_{in} density of insulating material

κ thermal diffusivity of insulation

θ boundary-layer momentum thickness, or angular position under the propellant grain star configuration (see Figure 10)

μ viscosity: μ_*, viscosity at the reference temperature, T_* in Sutherland's law (see Eq. 16)

ν kinematic viscosity, $\nu = \mu/\rho$

ρ local mass density of the gas at the edge of the boundary layer

ρ_s density of a particle

τ_W shear force at the wall

Subscripts

am arithmetic mean (wall-free stream)

aw adiabatic wall

i corresponding quantities in the related incompressible flow

in	virgin insulation material
0	stagnation conditions
TR	conditions at the point of transition from laminar to turbulent flow
∞	conditions at some reference point; e.g., in the free stream, at the propellant grain port, or at the transition point
*	conditions at the throat
total	length from start of aft closure to end of exit cone

REFERENCES

1. Pai, Shih-I, Fluid Dynamics of Jets, Van Nostrand Co., Inc., New York, 1954, p 139.

2. Price, F. C., et al. "Detail Design Optimization of Solid Propellant Rocket Motor Aft-Closure Nozzle Combinations (U)." Third Quarterly Progress Report (Confidential), Publication C-1635. Newport Beach, California: Aeronutronic Division, Ford Motor Company (April 5, 1962) (ASTIA AD 328912).

3. Grosh, R. J., "Transient Temperature in a Semi-Infinite, Porous Solid with Phase Change and Transpiration Effects." WADD TR 60-105. Lafayette, Indiana: Midwest Science Corporation (January 1960).

4. Scala, Sinclaire M. and Gilbert, Leon M. "Thermal Degradation of a Char-Forming Plastic During Hypersonic Flight." ARS Journal, Vol. 32 (June, 1962), pp. 917-924.

5. Carslaw, H. S., and Jaeger, J. C., Conduction of Heat in Solids. Second Edition. Oxford at the Claredon Press (1959), pp. 283-287.

6. Perry, J. H. (Editor), Chemical Engineer's Handbook. Third Edition. New York: McGraw-Hill Book Company (1950).

7. Hazelrigg, W. K. (Solid Rocket Plant, Aerojet-General Corp., Sacramento, California) Personal Communication (February 1962)(see Reference 41).

8. Solanski, A. H., et al. Dynamics of Separating Bodies Vol I "Theoretical Analysis," AFOSR TR-109, Holloman AFB, New Mexico; Directorate of Research Analysis, March 1961, p. 32 (ASTIA AD271998).

9. Nickelson, H. M. and Fields, J. P. "Some Experimental Techniques to Investigate the Mechanism of Flame Stabilization in the Wake of Bluff Bodies," Third Symposium on Combustion, Flame, and Explosion Phenomena Williams and Wilkins Co., 1949.

10. Bartz, D. R., "A Simple Equation for Rapid Estimation of Rocket Nozzle Convective Heat-Transfer Coefficients," Jet Propulsion, Volume 27, No. 1 (January 1957) pp. 49-51.

11. Welsh, W. E., Jr., and Witte, A. B., "A Comparison of Analytical and Experimental Local Heat Fluxes in Liquid-Propellant Rocket Thrust Chambers," Tech, Report No. 32-43, Jet Propulsion, Pasadena, California. (See also ASME Paper No. 61-AV-59.)

12. Eckert, E. R. G., "Survey of Boundary Layer Heat Transfer at High Velocities and High Temperatures." WADC TR 59-624 (April 1960).

13. Krieve, Walter F. and Mason, David M., "Heat Transfer in Reacting Systems: Heat Transfer to Nitrogen Dioxide Gas under Turbulent Pipe Flow Conditions, A.I.Ch.E. Journal, Vol. 7 (June 1961) pp. 277-281. See also: JPL Progress Report 20-366.

14. Richardson, J. L., et al., "Heat Transfer in Reacting Systems -- Heat Transfer to N_2O_4 Gas Flowing Normal to a Heated Cylinder," Chemical Engineering Science, Vol. 13 (February 1961) pp. 130-142.

15. Culick, F.C.C., and Hill, J.A.F., "A Turbulent Analog of the Stewartson-Illingworth Transformation," Journal of the Aeronautical Sciences, Vol. 25, No. 4 (April 1958), pp. 259-262.

16. Truckenbrodt, E., "A Method of Quadrature for Calculation of the Laminar and Turbulent Boundary Layer in Case of Plane and Rotationally Symmetrical Flow," NACA TM 1379 (1955).

17. Schlichting, Hermann, Boundary Layer Theory, 4th Ed., McGraw-Hill (1960) p. 574. Ibid., p. 190.

18. Rotta, J. "Näherungsverfahren zur Berechnung Turbulenter Grenzschichten unter Benutzung des Energiesatzes." Mitteilungen Nr. 8. Göttingen: Max-Planck - Institut für Strömungsforschung (1953). See also, Rotta, J. "On the Theory of the Turbulent Boundary Layer." NACA TM 1344 (1953); and Rotta, J. "Schubspannungsverteilung und Energiedissipation bei Turbulenten Grenzschichten." Ingenieur-Archiv, Vol. 20 (1952), pp. 195-207.

19. Michel, R., "Etude de la Transition sur les Profile D'Aile; Establishment d'un Critere le Determination du Point de Transition et Calcul de la Trainee de Profile en Incompressible," ONERA Report 1/1578A (1951). See also, ONERA (France) Publication No. 58 (1952).

20. Smith, A.M.O., "Transition, Pressure Gradient and Stability Theory," 9th International Congress of Applied Mechanics, Vol. 4, (1957) Brussels, Belgium: University of Brussels, pp. 234-244. See also, Smith, A.M.O. and Gamberoni, Nathalie "Transition, Pressure Gradient and Stability Theory." Report No. ES 26388. El Segundo, California: Douglas Aircraft Co., Inc. (Aug. 31, 1956).

21. Eckert, E., "Survey on Heat Transfer at High Speeds, " WADC Technical Report 54-70, Wright Air Development Center (April 1954) p. 37-39.

22. Rubesin, Morris W. "A Modified Reynolds Analogy for the Compressible Turbulent Boundary Layer on a Flat Plate." NACA TN 2917 (1953).

23. Ziebland, H. "Heat Transfer Problems in Rocket Motors," J. British Interplanetary Society, Vol. 14 (1955) pp. 249-264 (see page 257 in particular).

24. Bartz, Donald B. "An Approximate Solution of Compressible Turbulent Boundary Layer Development and Convective Heat Transfer in Convergent-Divergent Nozzles." External Publication No. 243. Pasadena, California: Jet Propulsion Laboratory (1955). See also, ASME Transactions (Nov., 1955), pp. 1235-45.

25. Johnson, H. A., and Rubesin, M. W., Trans, ASME, Vol. 71, pp. 447-456 (1949).

26. Ungar, E. W., "Particle Impacts on the Melt Layer of an Ablating Body," ARS Journal, Vol. 30 (September 1960) pp. 799-805.

27. Farbar, L., and Morley, M. J., "Heat Transfer to Flowing Gas-Solids Mixtures in a Circular Tube," Ind. and Eng. Chemistry, Vol. 49, No. 7 (July 1957) p. 1143.

28. Carlson, D. J., "Experimental Determination of Thermal Lag in Gas-Particle Nozzle Flows." Accepted for publication in ARS Journal (1962).

29. Stull, V. Robert, and Plass, Gilbert N., Emissivity of Dispersed Carbon Particles," Journal of Optical Society of America, Vol. 50, No. 2, pp. 121-129 (February 1960).

30. Van De Hulst, H. C., Light Scattering by Small Particles, Wiley, New York, 1957, p. 129.

31. Kliegal, J. R., and Nickerson, G. R., "Flow of Gas-Particle Mixtures in Axially Symmetric Nozzles," ARS Preprint 1713-61 (1961).

32. Spalding, D. B., "The Theory of Melting Ablation, with Vaporisation, Gas-Phase Chemical Reaction, Surface Pyrolysis, and Transient Effects," Journal of the Aeronautical Quarterly, London (August 1961), pp 237-74.

33. Adams, Mac C., "Recent Advances in Ablation," ARS Journal, Vol. 29 (September 1959) pp. 625-632.

34. Barriault, R. J. and Yos, J., "Analysis of the Ablation of Plastic Heat Shields that Form a Charred Surface Layer," ARS Journal, Vol. 30 (September 1960) pp. 823-829.

35. Scala, S. M., "A Study of Hypersonic Ablation," Space Sciences Laboratory, General Electric, Missile and Space Vehicle Department, Technical Information Series No. R59SD438 (September 1959) Presented at Tenth International Astronautical Federation Congress, 29 August - 5 September, 1959, London, England.

36. Munson, T. R., and Spindler, R. J., "Transient Thermal Behaviour of Decomposing Materials," Research and Advanced Development Division, Avco Corporation, Wilmington, Massachusetts (1961).

37. Lafazan, S. and Siegel, B., "Ablative Thrust Chambers for Space Application," Paper No. 5, Presented at American Institute of Chemical Engineers, 46th National Meeting, Los Angeles, Calif. (February 5, 1962). Los Angeles, California: Aerospace Corp.

38. Schwartz, H. S. (Editor), "Conference on Behavior of Plastics in Advanced Flight Vehicle Environments." WADD Technical Report 60-101, ASTIA AD 247100 (September 1960).

39. Brookley, C. E , "Measurements of Heat Flux to the Internal Insulation of a Rocket Motor." Allegany Ballistics Laboratory Research Progress Report A2-53. Cumberland, Maryland: Hercules Powder Company (November 2, 1961).

40. Baer, D., and Ambrosio, A., "Heat Conduction in a Semi-Infinite Slab with Sublimation at the Surface." STL report TR-59-0000-00610. Los Angeles, California: Space Technology Laboratories, Incorporated (February 24, 1959).

41. Epstein, G., "Status Report on Internal Insulation." Aerojet Report RD-R60-86. Azusa, California: Aerojet-General Corporation (September 26, 1960). Also; Hartz, Walter A., Personal Communication, General Tire and Rubber Co., Akron, Ohio (May, 1962).

42. Bradshaw, W. G., and Matthews, C. O., "Properties of Refractory Materials: Collected Data and References." LMSD-2466. Sunnyvale, California: Missile Systems Division, Lockheed Aircraft Corporation (June 24, 1958), ASTIA AD 205452, Page 73.

43. Smith, J. M., Introduction to Chemical Engineering Thermodynamics. New York: McGraw-Hill Book Company (1949), p. 94.

44. Jones, Warren C. (Materials Central, Wright-Patterson AFB, Ohio). Personal Communication (June 22, 1962).

45. R. Sehgal, "An Experimental Investigation of a Gas-Particle System." Technical Report No. 32-238. Pasadena, California: Jet Propulsion Laboratory (March 16, 1962).

NUCLEAR PROPULSION FOR ORBIT-BASED SPACECRAFT

C. J. Wang
Aerospace Corporation
Systems Research and Planning Division
Los Angeles 45, California

ABSTRACT

A cost analysis is made for nuclear and chemical vehicles for orbit-based spacecraft to show the significant cost advantage of nuclear propulsion. Conversion laws for equivalent nuclear and chemical components are introduced to establish equitable vehicle performance comparison as a basis for the cost study. Some design considerations for spacecraft, including such factors as vehicle size and thrust-to-weight ratio, are also examined.

NUCLEAR PROPULSION FOR ORBIT-BASED SPACECRAFT

C. J. Wang
Aerospace Corporation
Systems Research and Planning Division
Los Angeles 45, California

INTRODUCTION

Nuclear rocket propulsion has been of interest and the subject of study since the end of World War II. These studies have indicated that, although nuclear rockets have an abundant source of energy and potentially possess a high specific impulse, the high weight of the nuclear engine and the low density of hydrogen introduce a factor of structural inferiority. Because of this factor, it has been difficult to determine the areas in which the nuclear rocket would prove to be superior to the chemical rocket.

Historically, it was once believed that nuclear propulsion would prove superior to chemical propulsion in an ICBM of large payload and extreme range capability. However, studies indicated this to be untrue even for a payload of up to 100,000 pounds and a range of up to 8500 nautical miles. In addition, the nuclear rocket does not demonstrate any superiority over chemical rockets for use in the expendable space boosters. The advantage of lower gross weight of the nuclear system is cancelled by the attendant bulkiness, greater hardware weight, and objectionable nuclear-radiation effects.

In spite of these debits on the side of nuclear propulsion, the hard fact of the high cost of future space operations has stimulated new interest in nuclear rockets, as has the progress made by the NASA/AEC nuclear rocket development program.

Basically, the high specific impulse of a nuclear rocket represents a propellant-saving characteristic, and this saving increases rapidly with the velocity gain required of the vehicle. In the case of orbit-based spacecraft, high values of velocity gain will be essential and the unit price of propellant will be high due to the high cost of delivery to orbit. As a result, propellant cost is a prime consideration, and propellant saving derivable from nuclear propulsion would be an important advantage. This advantage would become even more significant in view of the overwhelming magnitude of the mission cost of future military space operations.

In this paper, therefore, nuclear rocket propulsion will be evaluated for use with orbit-based spacecraft in terms of its potential for reducing

mission cost. In this evaluation, nuclear vehicle design and performance is compared with that of chemical vehicles, and optimum design of nuclear vehicles is considered.

SPACE MISSIONS AND SPACECRAFT

Military orbit-based spacecraft may be used for offensive or defensive maneuvers, maintenance or rescue missions, command and control operations, or in a logistics capacity. Such missions would require velocity gain values on the order of 30,000 to 40,000 feet per second (Fig. 1). Requirements of this magnitude would be impracticable, to say the least, for a single-stage chemical vehicle, and multiple-stage vehicles would be undesirable from the standpoint of repeated use. It is for just such requirements that the nuclear vehicle has the advantage since it can easily provide these velocity increments in a single stage and, at the same time, maintain a reasonable payload capability.

A typical vehicle can be visualized in terms of the missions just mentioned. The duration of any of these missions could be months or longer; therefore, adequate shielding against solar flare radiations would have to be provided. Based on available data, it can be estimated that 70,000 pounds of shielding would be required to protect a three-man crew from one major flare, which will render the additional shielding requirements for a nuclear rocket negligible.

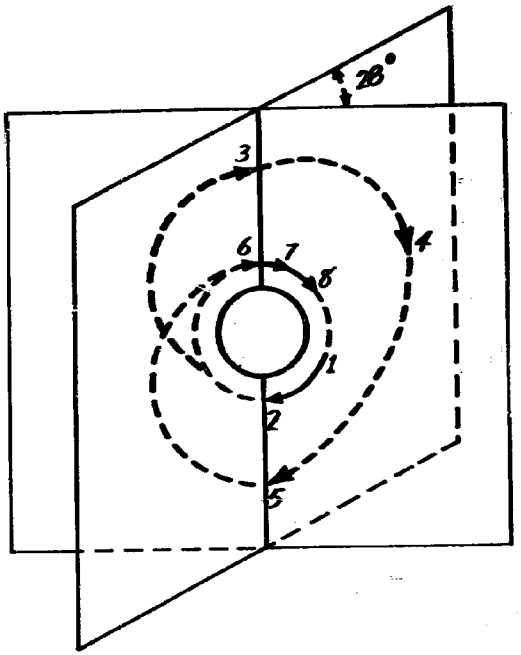
Should no environmental shielding be required for a mission, the shielding required will be specifically designed for the nuclear reactor radiations and will weigh approximately 10,000 pounds. These shielding weights and the requirements of a reasonable net payload have been used to estimate the range of some significant weights of a spacecraft. These estimates are shown on Fig. 2. From these weight figures, it must be concluded that the spacecraft to be considered is very large; for a velocity gain of 30,000 feet per second, the gross weight ranges from a low of about 160,000 pounds to a high of 800,000 pounds.

We now have some idea of the vehicle size; we have yet to consider the size of the engine. Since the spacecraft being considered is orbit-based, the problem of gravity is removed and the engine size is arbitrary; however, optimization is still desirable. Figure 3 illustrates the variation of gross-weight-to-payload ratio with the thrust-to-weight ratio. A large engine, that is, one with a high thrust-to-weight ratio, imposes a high structural factor on the vehicle and the gross weight-to-payload ratio increases. With a small engine, or low thrust-to-weight ratio, the inherent gravitational energy loss results in an increase in gross weight-to-payload ratio. An optimum engine size, which would result in a minimum gross weight-to-payload ratio, would correspond to a thrust-to-weight ratio in the range from 0.3 to 0.6.

NUCLEAR AND CHEMICAL VEHICLE PERFORMANCE COMPARISON

While nuclear spacecraft appears to be attractive, the reasons for the selection of nuclear propulsion over chemical propulsion for a spacecraft

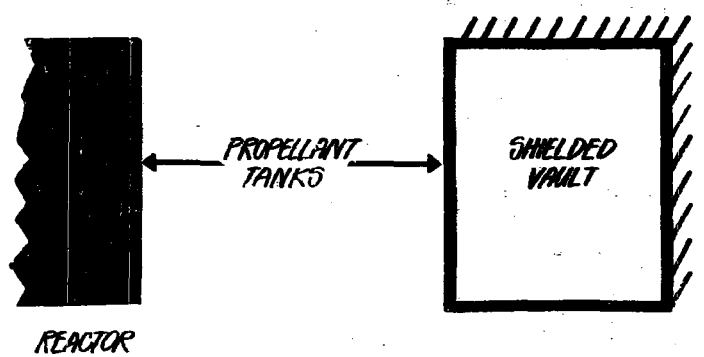
TYPICAL ACTION-SEQUENCE



	ΔV FT/SEC
(1-2) RENDEZVOUS WITH FUEL-STATION	2000
(2-3) ASCEND TO 24-HOUR ORBIT	8000
(3-4) INJECT INTO ORBIT AND CHANGE PLANE BY 28° INTO EQUATORIAL ORBIT	7000
(4-5) RENDEZVOUS IN 24-HOUR ORBIT	2000
(5-6) CHANGE PLANE 28° AND DE-ORBIT	7000
(6-7) INJECT INTO ORIGINAL 100 N.MI. ORBIT	8000
(7-8) RENDEZVOUS WITH CAPSULE	2000
TOTAL ΔV	36,000

Fig. 1. Spacecraft Velocity Gain Requirement

● RADIATION SHIELDING REQUIREMENT



SHIELD
CARBON,
LITHIUM HYDRIDE,
& LEAD

● WEIGHT ESTIMATES

ITEM	WEIGHT, LB	
	MINIMUM	MAXIMUM
RADIATION SHIELD	10,000 (REACTOR ONLY)	70,000 (REACTOR PLUS SOLAR FLARE)
USEFUL PAYLOAD	10,000	80,000
TOTAL	20,000	150,000
VEHICLE WT FOR $\Delta V \approx 30,000$ FT/SEC	160,000	800,000

Fig. 2. Orbit-Based Spacecraft Design Considerations

- LOW T/W_g IMPROVES VEHICLE STRUCTURAL FACTOR
- LOW T/W_g INTRODUCES ENERGY LOSS
- OPTIMUM T/W_g

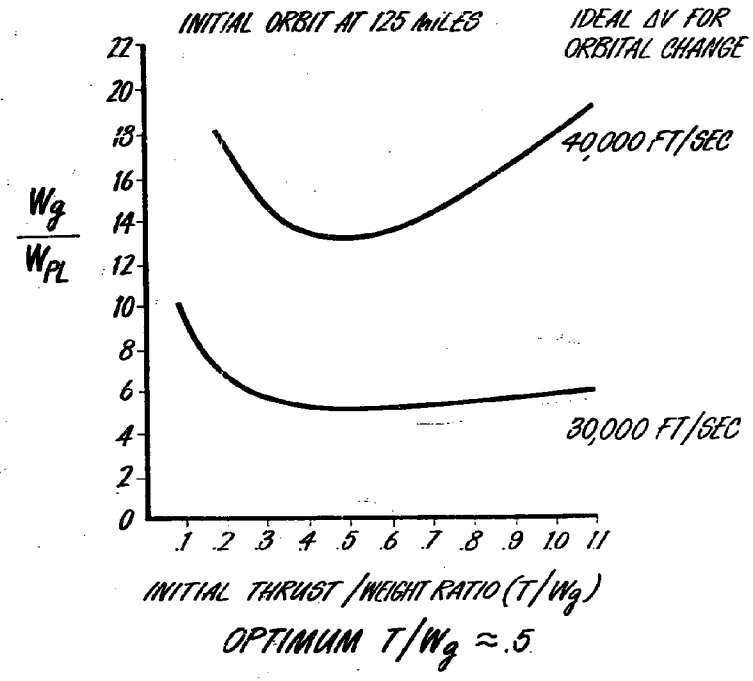


Fig. 3. Orbit-Based Spacecraft Design Considerations

must now be examined more exactly; comparisons must be made regarding the structural factor, the specific impulse, and the vehicle cost.

Previous comparative studies of nuclear and chemical vehicles have been controversial because different investigators have arbitrarily made inconsistent assumptions, mainly in the matter of weights. It is therefore important in the present investigation to ensure consistency of assumptions and to avoid any vagueness. A reasonable approach would be to establish precise weight conversion relationships between the corresponding nuclear and chemical system components. With the help of these component weight conversion laws, it will be possible to compare the performances of the nuclear and the chemical systems on a consistent and equitable basis. In case of any inconsistency, it will always be possible to trace any discrepancy to the responsible components, thus making it possible to resolve the difference. Furthermore, with this approach, both the nuclear and the chemical cases will reflect equally any optimism or pessimism that may have influenced assessment of the state-of-the-art.

To establish the component weight conversion laws, vehicle hardware weight is split into six parts; namely, engine, propellant tankage, thrust structure, pressurization system, control system, and miscellaneous components. Six weight ratios (K - values) corresponding to these six parts are defined below:

$$K_1 = \frac{\text{Engine Weight}}{\text{Vehicle Gross Weight}} \qquad K_4 = \frac{\text{Pressurization System}}{\text{Propellant}}$$

$$K_2 = \frac{\text{Tank Weight}}{\text{Propellant Weight}} \qquad K_5 = \frac{\text{Control System}}{\text{Gross}}$$

$$K_3 = \frac{\text{Thrust Structure}}{\text{Gross}} \qquad K_6 = \frac{\text{Miscellaneous}}{\text{Gross-Payload}}$$

For each of the six K-values, we derive a nuclear-chemical conversion law. It should be pointed out that the structural factor

$$\gamma = \frac{\text{Manufactured Weight}}{\text{Stage Weight}}$$

can be derived from the K-values by using the formula

$$\gamma = \frac{(K_1 + K_3 + K_5) + (1-e^{-\epsilon})(K_2 + K_4 + K_6)}{(K_1 + K_3 + K_5) + (1-e^{-\epsilon})(1 + K_2 + K_4)} \quad , \qquad \epsilon = \frac{\Delta V}{I_{spg}}$$

Conversion Between K_{1n} and K_{1c}

A nuclear engine consists of a nuclear reactor and some additional components that correspond essentially to those of a chemical engine. The

main components are the thrust chamber, turbo-machinery, and feed lines. For a given thrust, the thrust chamber weight varies inversely with the square root of chamber pressure. The weights of the turbo-machinery and the feed lines are proportional to the volume flow and the pressure involved. The engine thrust is usually designed to be proportional to the gross weight. The weight of a nuclear reactor is determined by the jet power, the thermal-to-jet-power conversion efficiency, and the reactor power density. Combining these weight considerations, we obtain the conversion law for K_1

$$K_{1n} = K_{1c}\left[f\left(\frac{p_c}{p_n}\right)^{1/2} + (1 - f)\left(\frac{p_n}{p_c}\right)\frac{\rho_c I_c}{\rho_n I_n}\right] + [2.18(10^{-5})]\left(\frac{a_o I_{sp}}{\eta P}\right)$$

where

a_o = initial thrust to weight ratio of vehicle

η = thermal to jet power conversion efficiency

P = reactor power density

f = fraction of chemical engine weight due to thrust chamber

p = thrust chamber pressure

ρ = liquid propellant density

Subscripts c and n indicate chemical and nuclear, respectively.

Conversion between K_{2n} and K_{2c}

The weight of a propellant tank is proportional to the tank volume and the tank pressure. Assuming the nuclear tank to consist of a cylindrical body with spherical head and the chemical tank to have the same configuration plus a spherical partition, we obtain the conversion law of K_2

$$K_{2n} = \frac{L/D}{L/D + 1/2}\,(K_{2c})\left(\frac{\rho_c}{\rho_n}\right)$$

where L/D is the tank length to diameter ratio, and ρ is the propellant density. It should be mentioned that this formula has neglected the hydrostatic pressure effect on the tank thickness--an omission acceptable for the vehicle sizes and acceleration considered here. Also, the partition wall in the chemical tank, including the insulation, is assumed to have the same weight per unit area as the outer wall.

Conversion Laws for K_3, K_4, K_5, K_6

The four remaining K-values are unaffected by the type of engine, whether nuclear or chemical. Thus, the conversion laws are simply:

$$K_{3n} = K_{3c}$$

$$K_{4n} = K_{4c}$$

$$K_{5n} = K_{5c}$$

$$K_{6n} = K_{6c}$$

As mentioned earlier, the structural factor can be derived directly from K-values. Figure 4 gives corresponding nuclear and chemical structural factors derived from two corresponding sets of K-values. These K-values are obtained for orbit-based spacecraft. The specific impulse is assumed to be 150 feet per second for the nuclear engine, and 450 feet per second for the chemical engine. These are considered corresponding values because they are obtainable under equivalent conditions.

At this point it is a simple matter to compute the vehicle performance in terms of W_g/W_{pl}, W_p/W_{pl}, and W_m/W_{pl} versus ΔV, where

W_{pl} = payload weight

W_g = gross vehicle weight

W_p = propellant weight

W_m = vehicle hardware weight

The resulting performance curves are given in Fig. 5.

It is of interest that, for a given payload, the gross weight and propellant weight of a nuclear vehicle are less than the corresponding weight of a chemical vehicle. However, up to a velocity of approximately 35,000 feet per second, the nuclear vehicle has a greater manufactured weight than its chemical counterpart.

VEHICLE COST ANALYSIS

The performance curves given in Fig. 5 are used to determine pertinent vehicle design parameters for analysis of the cost of nuclear and of chemical spacecraft. For any given payload and velocity gain, the performance curves will give the vehicle gross weight, propellant weight, and manufactured weight. The engine thrust, obtained by multiplying the gross weight

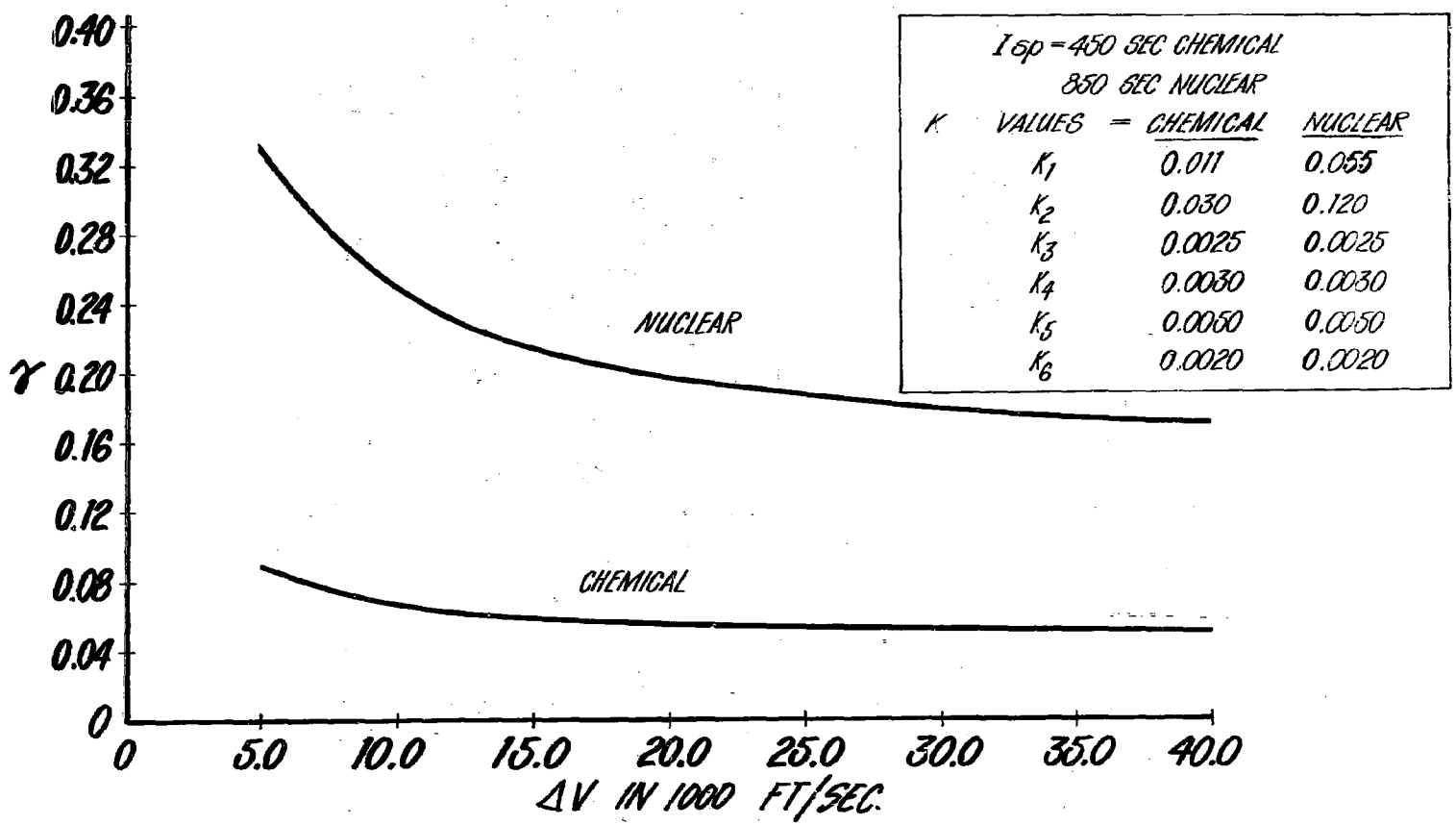


Fig. 4. Nuclear and Chemical Structural Factor Vs. Velocity Gain

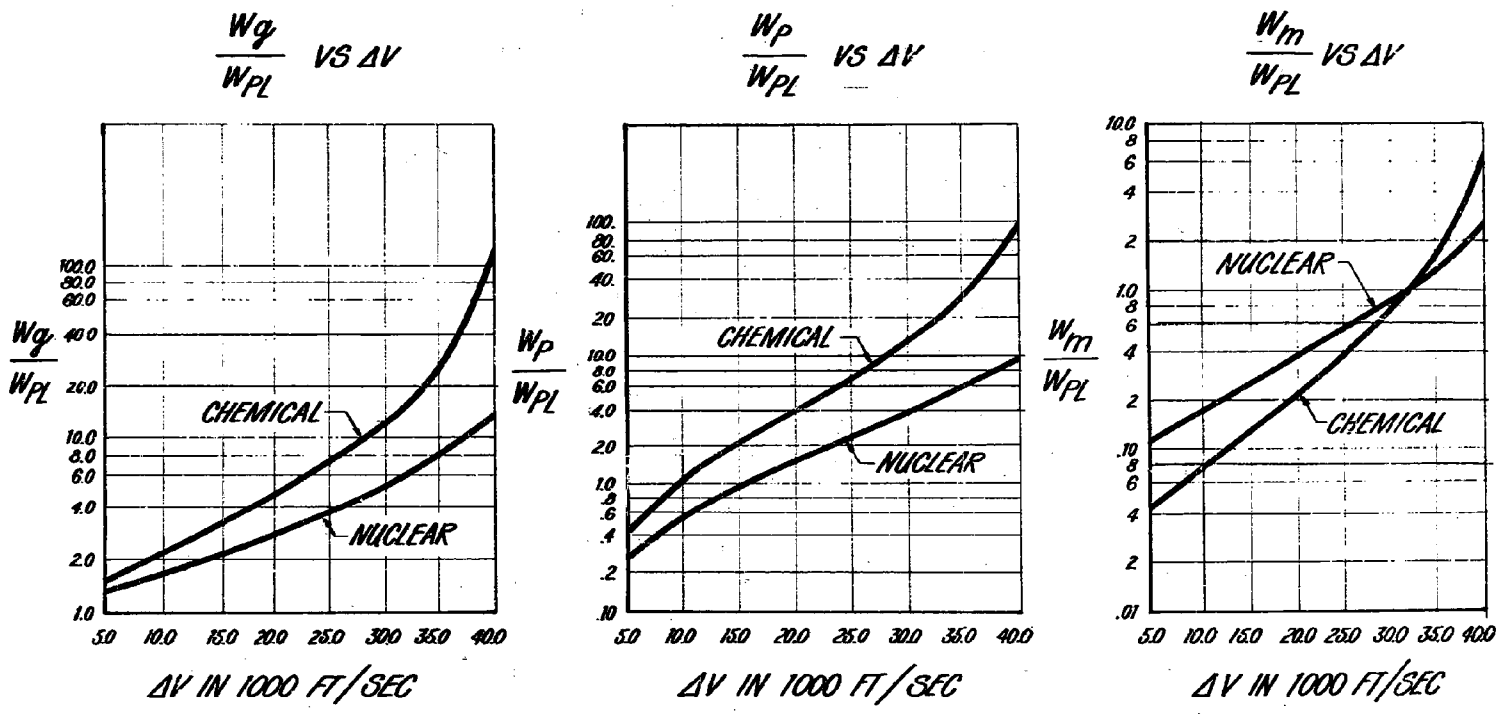


Fig. 5. Comparative Performance of Single-Stage Nuclear and Chemical Vehicles

by 0.5, will determine the engine size. The major design parameters obtained by this process for three vehicles are listed in Table 1.

These design parameters may now be used for costing the vehicle. The assumptions to be made in this costing process are listed in Table 2. These assumptions, except for the cost of delivery to parking orbit, are essentially conventional. Two assumptions are made for parking-orbit delivery costs: a projected figure of $20 per pound, as well as the more conventional figure of $200 per pound.

Figure 6 shows the results of a cost analysis for a large number of nuclear spacecraft. The cost in dollars per pound of payload is plotted against velocity gain. The cost of delivery to a parking orbit is assumed to be $200 per pound. Three curves are shown, corresponding to three trip numbers over which the hardware cost is amortized. Thus, the top curve, for $n = 1$, gives the total cost of propellant and hardware, assuming that the vehicle will be used only once. With increasing n, the cost curve is lowered. It must be noted, however, that for amortization trips in excess of ten the additional gain is negligible.

Figure 7 shows a similar set of curves for an assumed cost of delivery to a parking orbit of $20 per pound. As in the previous case, the additional gain is very slight after the tenth amortization trip.

To illustrate the magnitude of the cost of a typical mission, Fig. 8 has been prepared for $\Delta V = 30,000$ feet per second for various vehicle sizes. The hardware cost is amortized over ten trips. The parking-orbit delivery cost is $200 per pound. For vehicle gross weight of 800,000 pounds, the mission cost is 120 million dollars. Even for a gross weight of 200,000 pounds, the mission cost is still as much as 30 million dollars -- staggering cost figures.

To determine how the cost of the nuclear vehicle will compare with the cost of the corresponding chemical vehicle, similar calculations were also made for a large number of chemical vehicles. The results of these calculations are shown in Fig. 9 along with the nuclear vehicle cost for $n = 10$, and for both values of the delivery cost. Suppose we consider the $200 per pound curves. It is seen that the chemical vehicle is more expensive than the nuclear vehicle by a factor of at least three and possibly as much as ten, depending on the value of ΔV. The same observation can be made about the $20 per pound curves.

These results indicate that, whereas military space operations will inevitably be expensive, they will be even more expensive if chemical rather than nuclear rockets are used. For a mission involving a large number of vehicles or vehicles having many repeated uses (and thus a high mission cost), the difference between the costs of nuclear and chemical rockets may mean the difference between possibility and impossibility for the mission.

Table 1. Sample Nuclear-Vehicle Parameters

PARAMETERS	VEHICLES		
	I	II	III
Velocity Increment, Ft/sec	30,000	30,000	30,000
Payload, Lb	20,000	69,000	150,000
Gross Weight, Lb	160,000	400,000	800,000
Engine Thrust, Lb	80,000	200,000	400,000
Specific Impulse, Sec	850	850	850
Propellant Weight, Lb	110,300	275,000	550,000
Burnout Weight, Lb	49,700	125,000	250,000
Vehicle Structure Weight, Lb	29,700	56,000	100,000

Table 2. Cost Assumptions

ITEM	COST (DOLLARS)
Engine:	
—Chemical	1.00 per lb thrust
—Nuclear, Hardware	150.00 per lb weight
Fuel Elements	200.00
Vehicle Structure:	70.00 per lb
Propellants:	
H^2	0.25 per lb
Lox	0.03 per lb
Delivery to low orbit cost:	
(A)	200.00 per lb
(B)	20.00 per lb

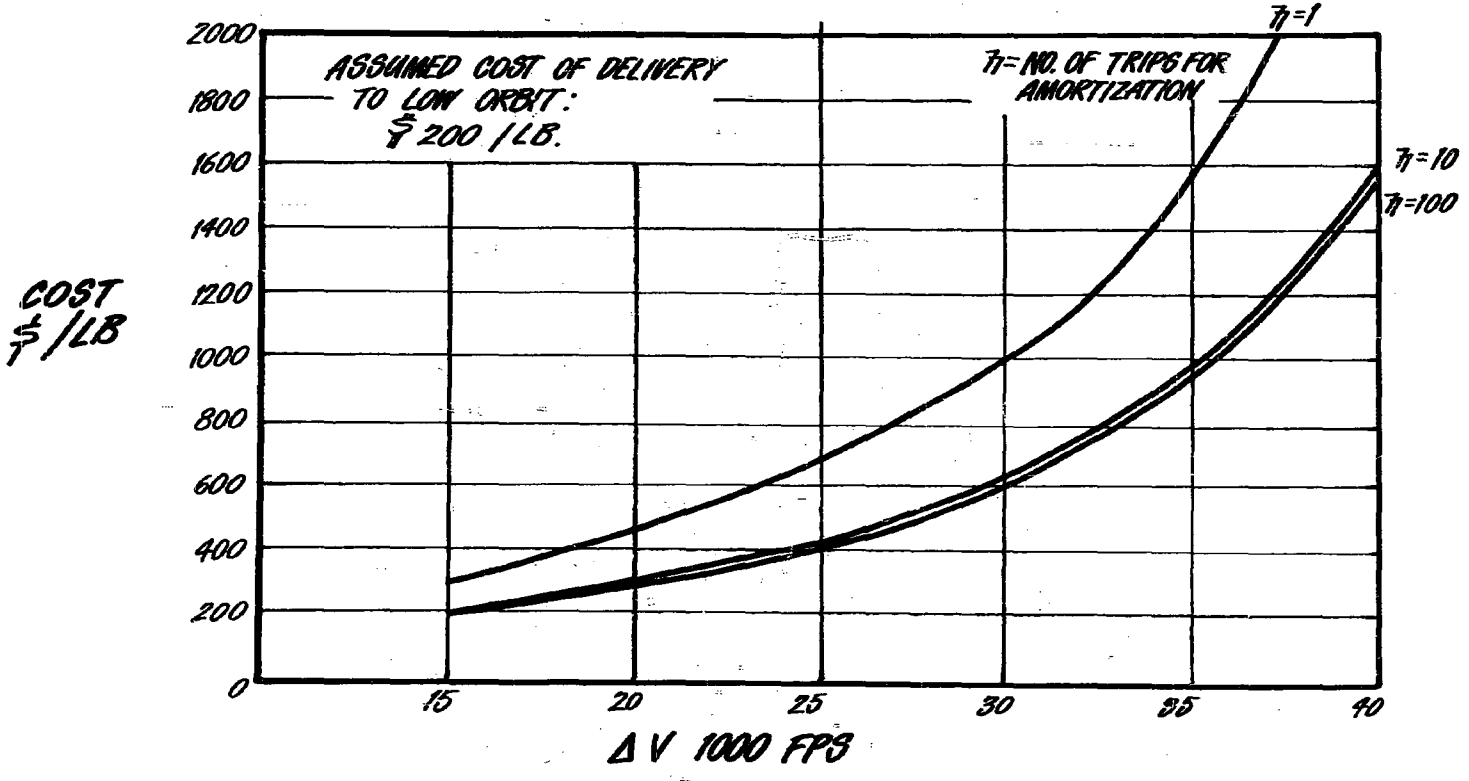


Fig. 6. Nuclear Spacecraft--Cost per Pound of Payload Vs. Velocity Gain (cost includes propellant and hardware amortization)

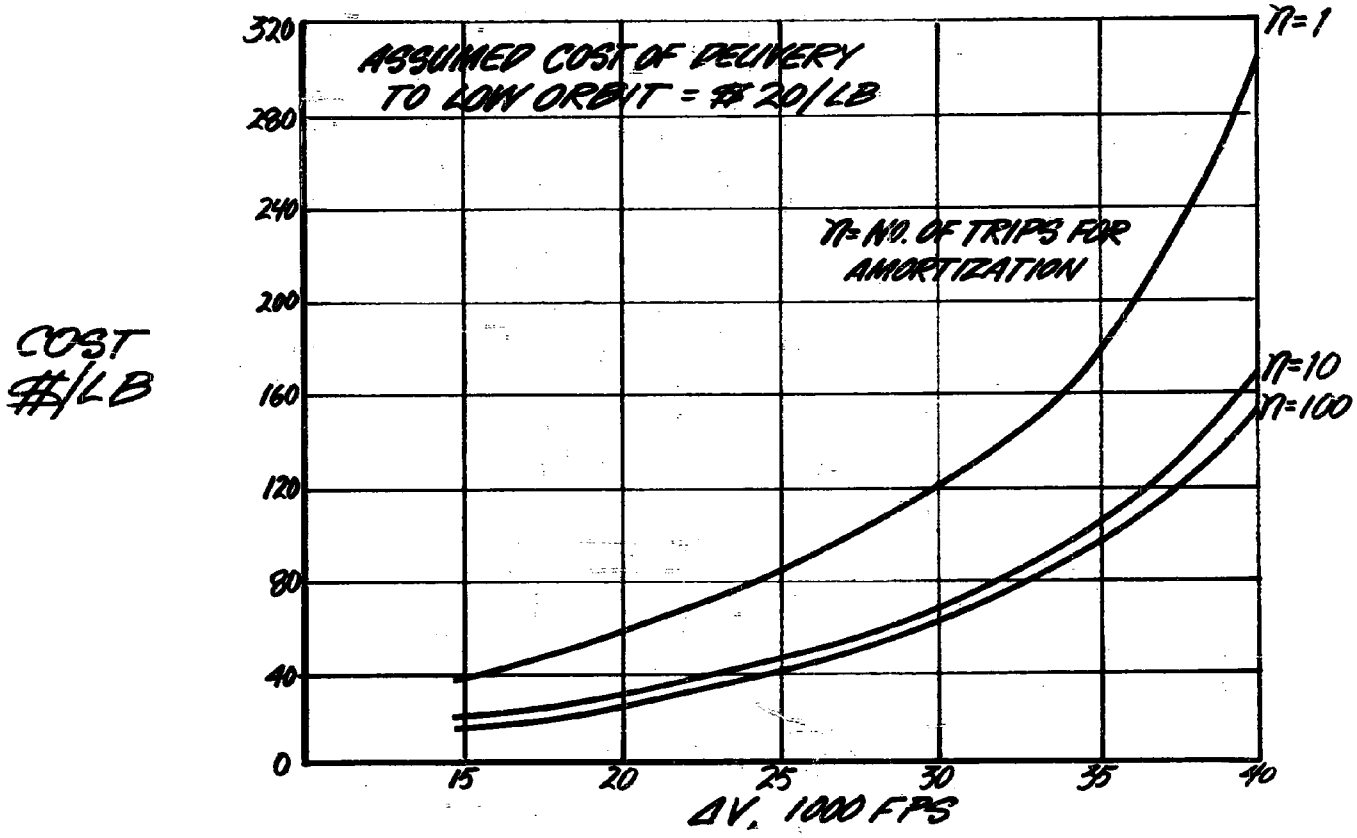


Fig. 7. Nuclear Spacecraft--Cost per Pound of Payload Vs. Velocity Gain (cost includes propellant and hardware amortization)

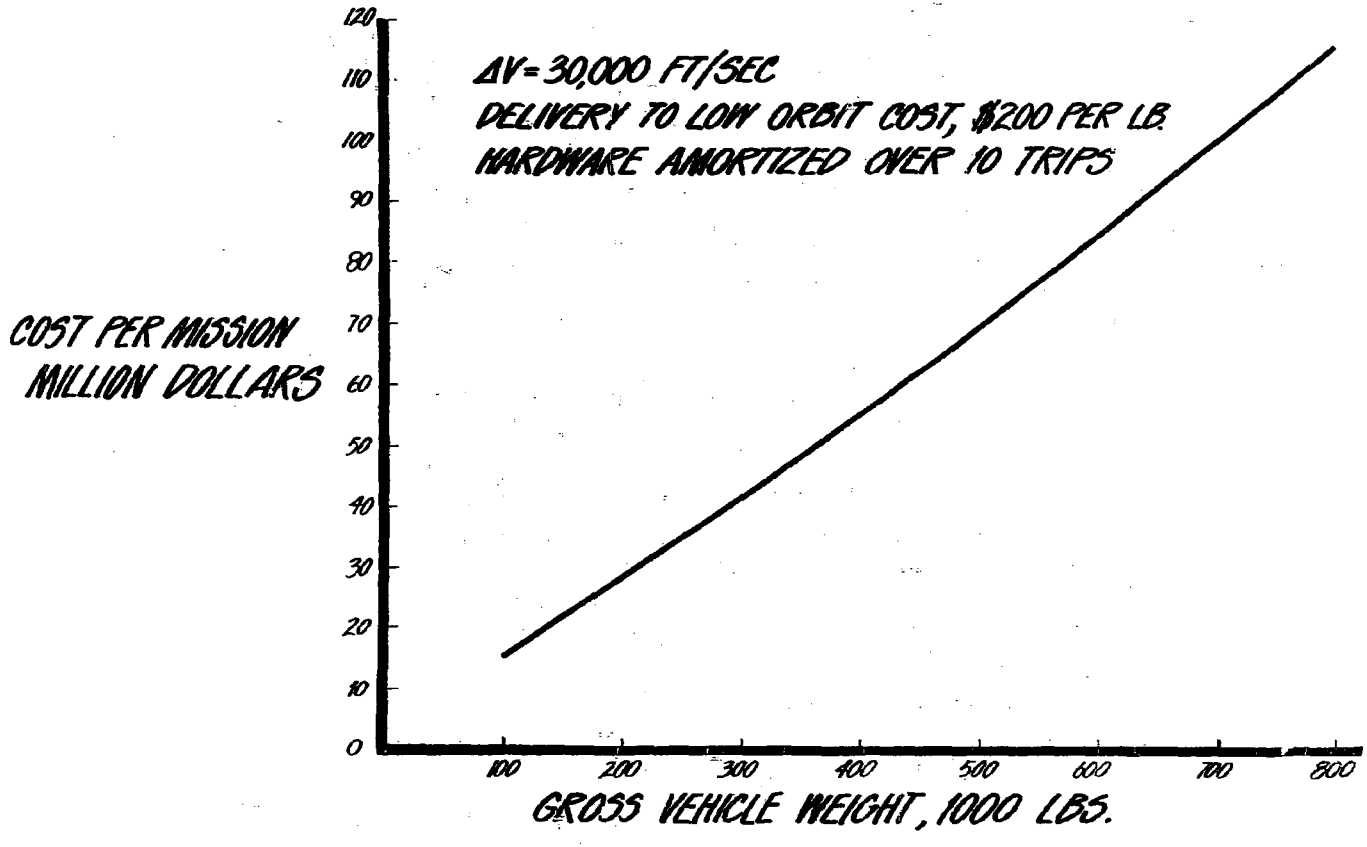


Fig. 8. Nuclear Orbit-Based Spacecraft Mission Cost

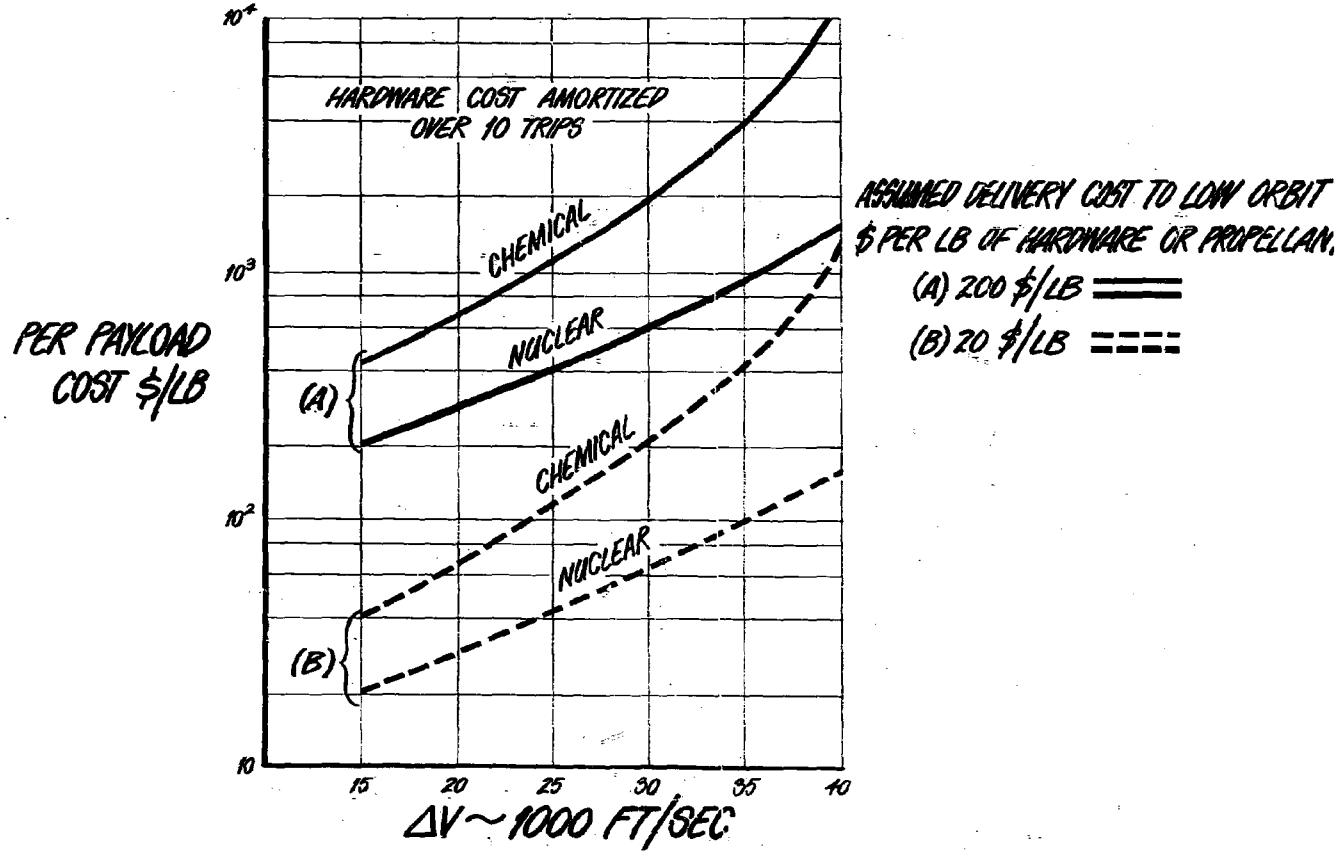


Fig. 9. Nuclear and Chemical Orbit-Based Spacecraft--Cost in Dollars per Pound of Payload Vs. Velocity Gain

THE SMALL ENGINE

The value of a small engine has been mentioned recently in connection with nuclear spacecraft. It might be suggested that an engine on the order of, say, 10,000 pounds thrust might be considered in place of the large engines considered thus far. It may be argued that it will cost less to develop a smaller engine. However, there are basic problems connected with a small engine that must not be overlooked.

As shown in Fig. 10, a small engine used in conjunction with a large vehicle will result in energy loss due to the small thrust-to-weight ratio. On the other hand, a small engine used in conjunction with a small vehicle will be penalized by a high vehicle structural factor. The effects of energy loss and increased structural factor on vehicle cost are shown in Fig. 11. It can be seen that these effects could wipe out the potential cost advantage of the nuclear rocket.

CONCLUSIONS

Based on the considerations presented in this paper, it may be concluded that use of nuclear rocket propulsion for orbit-based spacecraft could significantly reduce the high cost of future military space missions. The nuclear spacecraft should be of large size, should have a thrust-to-weight ratio of approximately 0.5, and should be re-usable for about ten times. As just shown, the use of a small nuclear engine would make it impossible to realize the potential cost advantages of nuclear propulsion for orbit-based spacecraft.

ACKNOWLEDGEMENT

The author wishes to acknowledge valuable suggestions by Mr. A. F. Donovan on the method of approach of the paper. He is also indebted to Dr. A. Mager and Mr. J. H. Irving for helpful discussions on the component weight scaling laws and military space missions. The assistance rendered by Mr. M. J. Russi and Miss Mary E. Travis during the preparation of the paper is gratefully acknowledged.

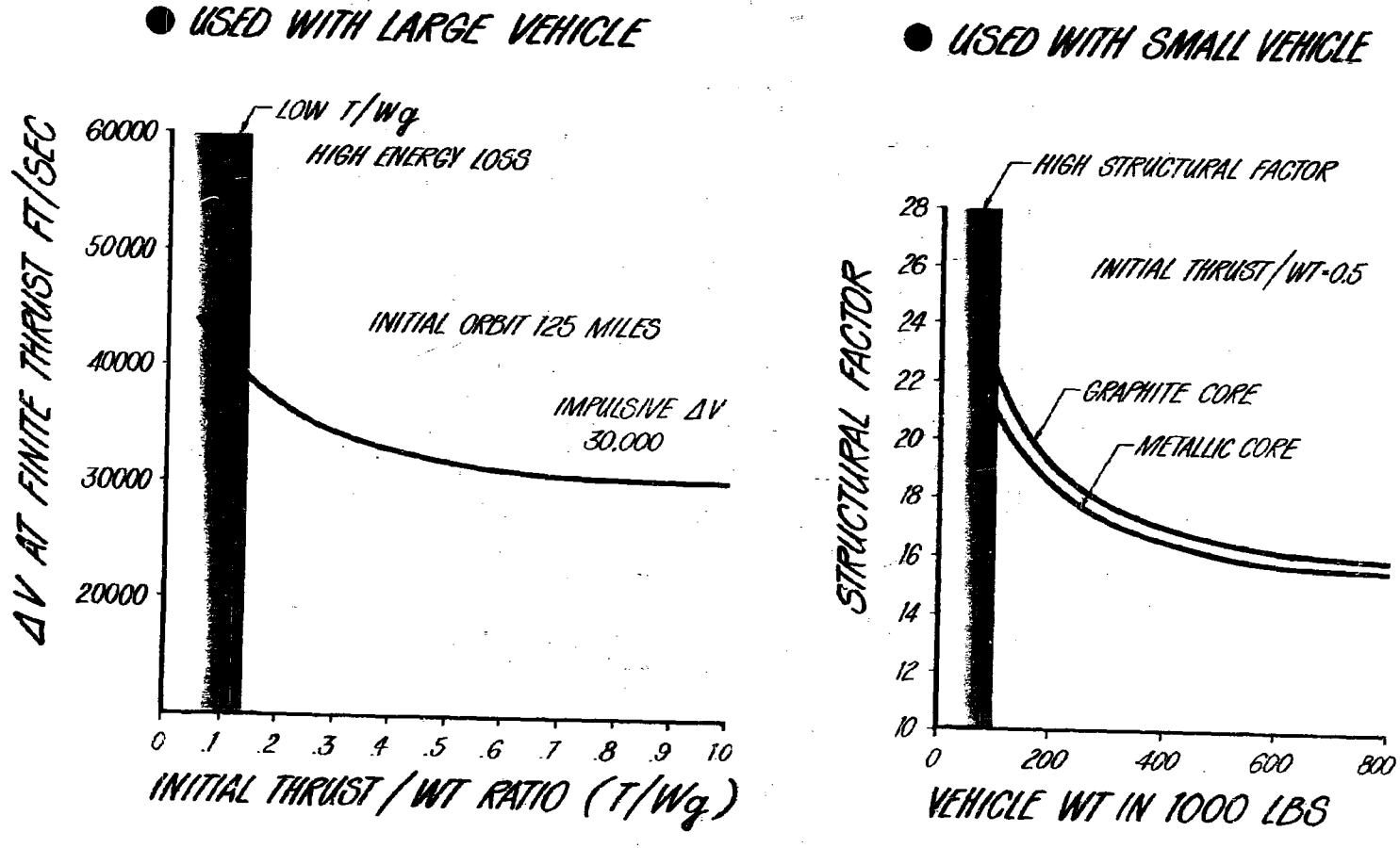


Fig. 10. Effects of Small Engine used with Both a Small and a Large Vehicle

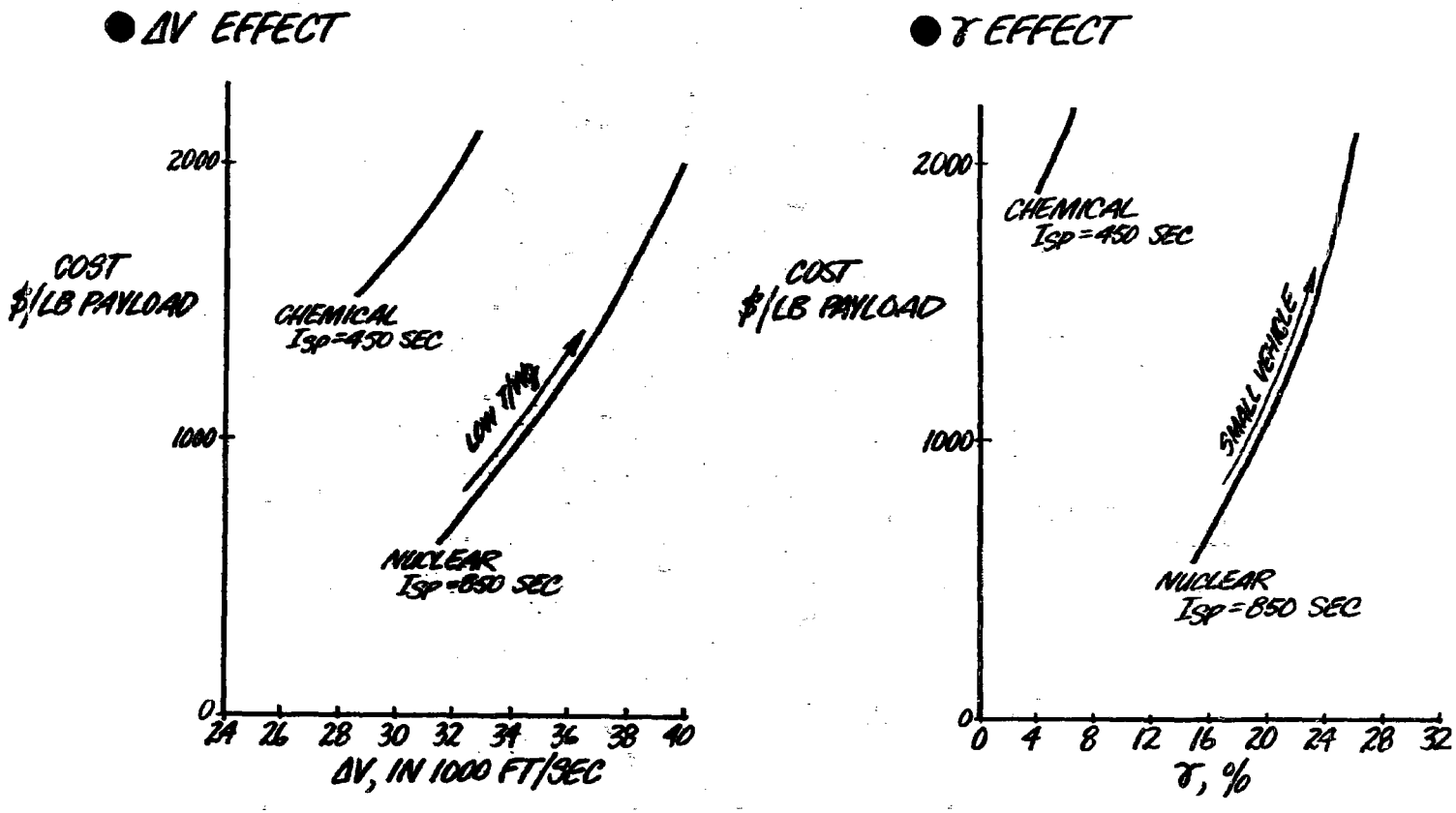


Fig. 11. Effects of the Use of a Small Engine on Vehicle Cost

STRUCTURES AND MATERIALS

MISSILE DESIGN FOR THE EFFECTS OF WINDS ALOFT

D. C. Bakeman

Aerospace Corporation
Engineering Division
Los Angeles 45, California

ABSTRACT

A major problem in the design of large booster missiles is the proper consideration of the effects from winds aloft. Since wind magnitude and direction are statistical in nature, the problem can only be solved by the use of statistical techniques.

This report suggests that the design requirement be a given probability of failure or launch delay due to winds aloft, or an optimization between cost (in terms of weight, schedules, etc.) and the probability of failure or launch delay. A methodology is then presented, along with pertinent background information and discussion, for designing to any of those requirements. Also presented is a procedure for a prelaunch wind check in which wind effects magnitudes are predicted for a flight.

The work described in this paper was performed under Air Force Contract 04(695)-69.

MISSILE DESIGN FOR THE EFFECTS OF WINDS ALOFT

D. C. Bakeman
Aerospace Corporation
Engineering Division
Los Angeles 45, California

1.0 INTRODUCTION

The problem of designing a large ground launched booster missile for the effects resulting from winds aloft (the effects of ground winds will not be considered) is an important and current problem in the missile field. Its importance is a result of the fact that, when a boost vehicle climbs through the atmosphere, winds can produce effects of sufficient magnitude to effectively "design" the vehicle in some areas. In other words, some areas of the overall missile system would not require as great a capability if it were not for winds aloft. It is possible to design a missile to withstand a wind criteria so severe that it will essentially never fail due to winds. However, by applying such overconservatism to the design of a booster for space vehicles, the required booster capabilities may be seriously increased or the payload-carrying capability seriously decreased.

The solution to this design problem requires the use of statistical techniques because the wind magnitude and velocity are statistical in nature. The various design requirements that this report considers are (1) a given probability of failure due to winds aloft, (2) a given probability of launch delay, if a launch is to be allowed only when the predicted wind effects are allowable; or (3) an optimum trade-off between cost (in terms of weight, complexity, etc.) and probability of failure or launch delay.

The solution provided in this report is a methodology which computes the probability of failure or launch delay for any given design. Therefore, to find a design that meets the given probability requirements, or to optimize the design, a cut-and-try process is required. The methodology could be termed an engineering approach - that is, it utilizes a number of assumptions in order to maintain practicality. These assumptions are presently necessary because they simplify computations, and because much of the statistical data used is limited in both quantity and accuracy. Efforts are being made to obtain better data, but major improvements will take years. The most unique aspect of the methodology is that it considers the combined effects of all random factors that can lead to failure (i.e., wind profiles, gusts, buffeting, in-spec missile parameter variations, etc.).

Some background information on winds, and a discussion of some current design techniques, will be provided to aid in understanding the methodology and placing it in perspective. Presentation of the methodology will

be more graphical than mathematical. This will hopefully increase clarity, although at an admitted sacrifice of mathematical rigor. Relevant information regarding design techniques for an unsymmetrical vehicle, and a prelaunch wind check procedure are placed in the Appendix to avoid a break in continuity.

2.0 BACKGROUND INFORMATION ON WINDS

Information regarding wind velocity versus altitude is divided into two parts which will be termed the wind profile, and the profile detail. The reason for this dichotomy is that balloon soundings have been used to obtain the greatest mass of wind data; and, since a standard balloon sounding produces average wind velocities over altitude layers of 2000 feet, these soundings do not produce complete data on gusts and shears. The wind information produced by a balloon sounding will therefore be termed wind profile, and that undetected by the balloon sounding will be termed profile detail. These are discussed separately. The concept of a wind rose is also discussed, since it is often encountered in wind studies and criteria.

2.1 Wind Profile

A typical wind velocity profile is shown in Figure 1. It can be termed "typical" because the velocity peaks in the altitude region where high winds are most probable. Actually, winds tend to peak again in the region of 200,000 feet, but air density there is so low that the wind effects on the missile are not normally important. To provide the reader with a "feel" for wind speeds: In the region of 35,000 feet at Cape Canaveral a wind velocity of 250 ft/sec (170 mph) is exceeded about 5 percent of the time during the month of the heaviest winds.

Unfortunately, the region of high winds is also the region in which most missiles experience the highest dynamic pressure, q*, even without winds. Thus, this region usually contains the critical altitude - that is, the altitude at which the probability of failure due to winds is greatest.

The magnitude of the effects caused by winds are a function of the history of the wind shears (the rate of change of velocity with altitude) and direction, as well as instantaneous wind velocity. A change in any of these parameters will change the magnitude history of the wind effects. Therefore, the only accurate way of computing the maximum wind effect magnitudes due to any given profile is to essentially simulate the flight of the missile through that profile.

2.2 Profile Detail

As previously explained, the profile detail consists of variations in wind velocity versus altitude that cannot be detected by a balloon sounding (see Figure 1). How much of this detail is shear and how much can be termed gusts is an academic question for the purpose of this report.

---

*$q = \frac{1}{2} \rho V^2$, where ρ is air density and V is velocity relative to the air mass.

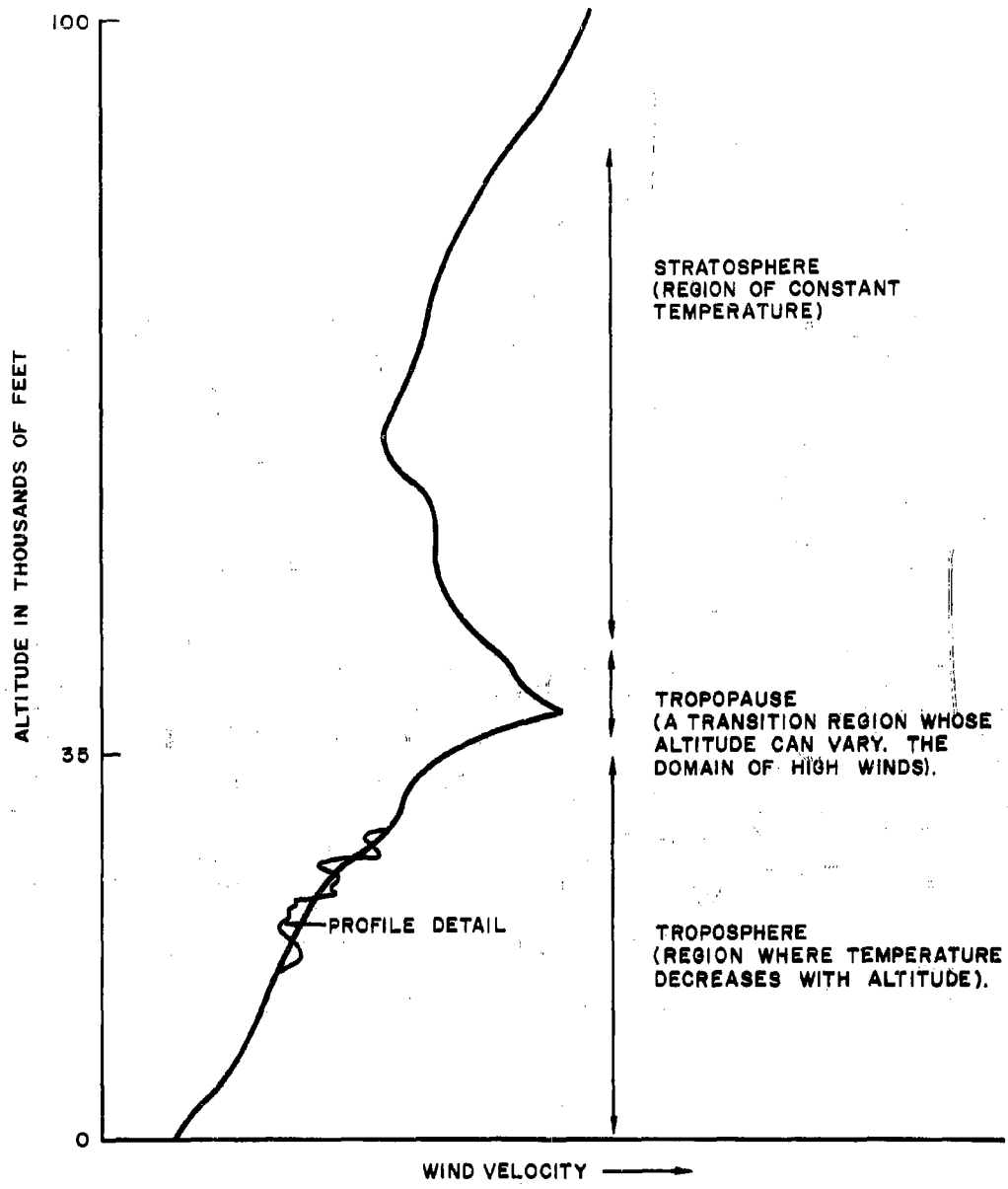


Figure 1. Typical Wind Velocity Profile

The importance of the profile detail is that it produces almost all of the dynamic effects (e.g., dynamic rigid and flexible body bending moments) as opposed to the quasi-steady state effects produced by wind profiles. Unfortunately, although the effects of the profile detail can be very significant, little useful and reliable statistical data presently exists regarding profile detail. The data that is available was obtained from acceleration instrumentation on aircraft during level flight, (7). This is, therefore, gust data and does not include shear effects that are not measured by a balloon, but will be encountered by a vertically rising missile.

2.3 Wind Rose

A wind rose is not used in this report, but an understanding of it is important because it involves several concepts commonly used in wind studies and wind criteria. A typical rose of magnitude and one of direction are shown in Figure 2.

The rose of Figure 2a plots wind magnitude versus compass direction at some altitude. The magnitude in each direction is usually the wind velocity that is equaled or exceeded some percent of the time assuming that the wind is in that direction. For example, one could find the 1 percent (probability of exceedance = 1 percent) wind velocity in a compass sector by considering all the winds in that sector at the altitude of interest, and computing the wind velocity that was equaled or exceeded 1 percent of the time. By doing this in each sector, a 1 percent magnitude rose can be constructed. One could also construct a magnitude rose by plotting the magnitude of wind component exceeded in each direction some percent of the time. The probability distribution of such a wind component magnitude in a given direction is found by breaking all wind vectors down into two components, one of which is in the given direction. This wind component rose would be of a different shape and have a different meaning than the previous rose where only total wind vectors were considered.

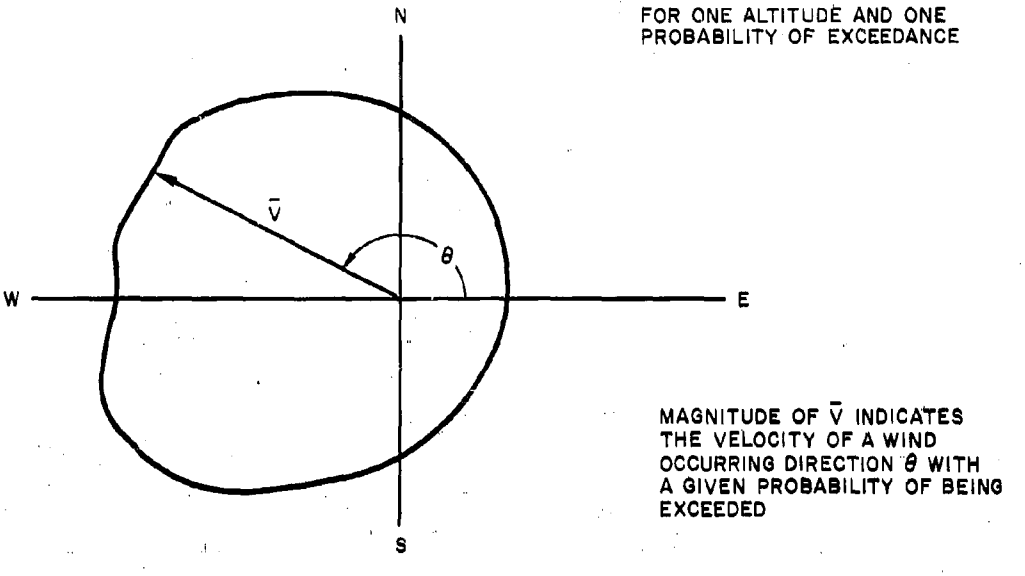
The rose of Figure 2b plots the probability of wind occurrence versus compass direction at some altitude. The distance from the origin to any point on the rose allows one to compute the probability of the wind occurring in any sector of the compass.

3.0 DEFINITION OF TERMS

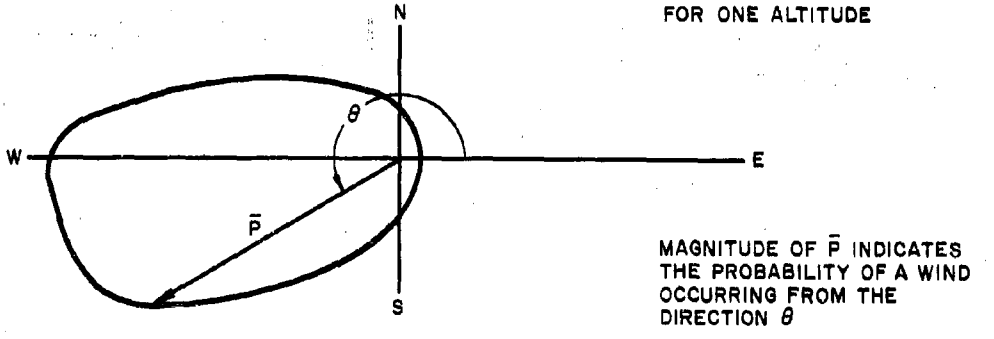
A number of important terms are used in this report. The terms wind profile and profile detail were defined in Section 2. Others of equal importance are defined below.

3.1 Wind Profile Effects And Effect Parameters

When a booster missile climbs through the atmosphere, the presence of a wind profile will cause a change (normally an increase) in aerodynamic forces on the missile. These forces produce a number of effects that will be termed wind profile effects. The most important of these effects to the present-day designer are listed below along with some of the effect parameters.



2a



2b

Figure 2. Typical Wind Roses

(1) Structural Loads

Effect parameters are: bending moment, etc.

(2) Control System Perturbations

Effect parameters are: thrust vector angle (that is, angle from effective null), integrated thrust vector angle (important for secondary injection thrust vector control), autopilot signals (e.g., gyro outputs), etc.

(3) Trajectory Perturbations

Effect parameters are: velocity vector, missile position, dynamic pressure, angle-of-attack, etc. Many of particular interest at staging or at guidance initiation.

3.2 Non-Wind Profile Effects and Factors

The effects of wind profiles were discussed above and some effect parameters were listed. These same effects can result from factors other than a wind profile, and the same parameters will be influenced. When these effects are caused by factors other than wind profiles, they will be termed non-wind profile effects. Some of the important non-wind profile factors that must usually be considered are:

(1) Profile Detail

Profile detail produces aerodynamic loads just as the wind profile does. However, its statistics are separately defined. Also, it tends to produce dynamics, while a wind profile changes so slowly that its effects are mainly quasi-steady state. The most important effect from profile detail is usually structural loads.

(2) Buffeting

Like profile detail, buffeting produces relatively rapid variations of aerodynamic loads with structural loads constituting the main effect.

(3) Propellant Sloshing

It will produce control system perturbations and structural loads.

(4) No-Wind Trajectory Perturbations

These result from variations (within specifications) of missile parameters such as thrust, pitch program commands, gyro drift, etc. They can produce aerodynamic loads and, as a result, all of the wind profile effects listed in Section 3.1.

Non-wind profile effects are important because the failure of a missile due to winds occurs when the magnitude of some effect parameter exceeds

its design allowable, and the magnitude of the parameter may be influenced by many factors. Thus, the computation of a probability of failure requires consideration of uncertainties due to all factors, both wind profile and non-wind profile.

The design procedures presented in this report use the artificial division of wind profile effects and non-wind effects when considering the magnitude of any effect parameter. The main reason for this is that wind profile effects are usually dominant and can be computed with much more accuracy than can non-wind profile effects.

3.3 Design Allowable

A design allowable is the magnitude of a missile system design parameter which is considered to constitute a flight failure when exceeded in flight.

3.4 Probability of Failure Due to Winds

The probability of failure due to winds aloft as used in this report will mean the probability that the magnitude of an effect parameter (see 3.2) will exceed its design allowable due to all factors when (1) all parts of the missile meet specifications, and (2) the missile is launched at a random time, within a defined period of the year, and from a given location.

3.5 Launch Allowables

The launching of large boosters carrying space vehicles are usually permitted only when the magnitudes of effect parameters that are predicted for the flight are below preassigned values. These preassigned values are termed launch allowables and are usually provided as a function of time-of-flight.

3.6 Probability of Launch Delay

The probability of launch delay due to winds aloft will mean the probability that a predicted magnitude of the effect parameters will exceed its launch allowable at some time-of-flight when (1) the launch is scheduled to occur at a random time, during a defined period of the year, and from a given location, and (2) the launch delay decision is always made at a given point in the countdown.

3.7 Probability of Exceedance

The probability of exceedance is the probability that a certain magnitude will be equalled or exceeded.

4.0 DESIGN REQUIREMENTS

When the designer is given the problem of designing a new missile, he must be given some requirements in regard to the effects of winds aloft. These requirements could specify (1) a probability of success in flying through winds when a launch is made without any consideration of winds, or

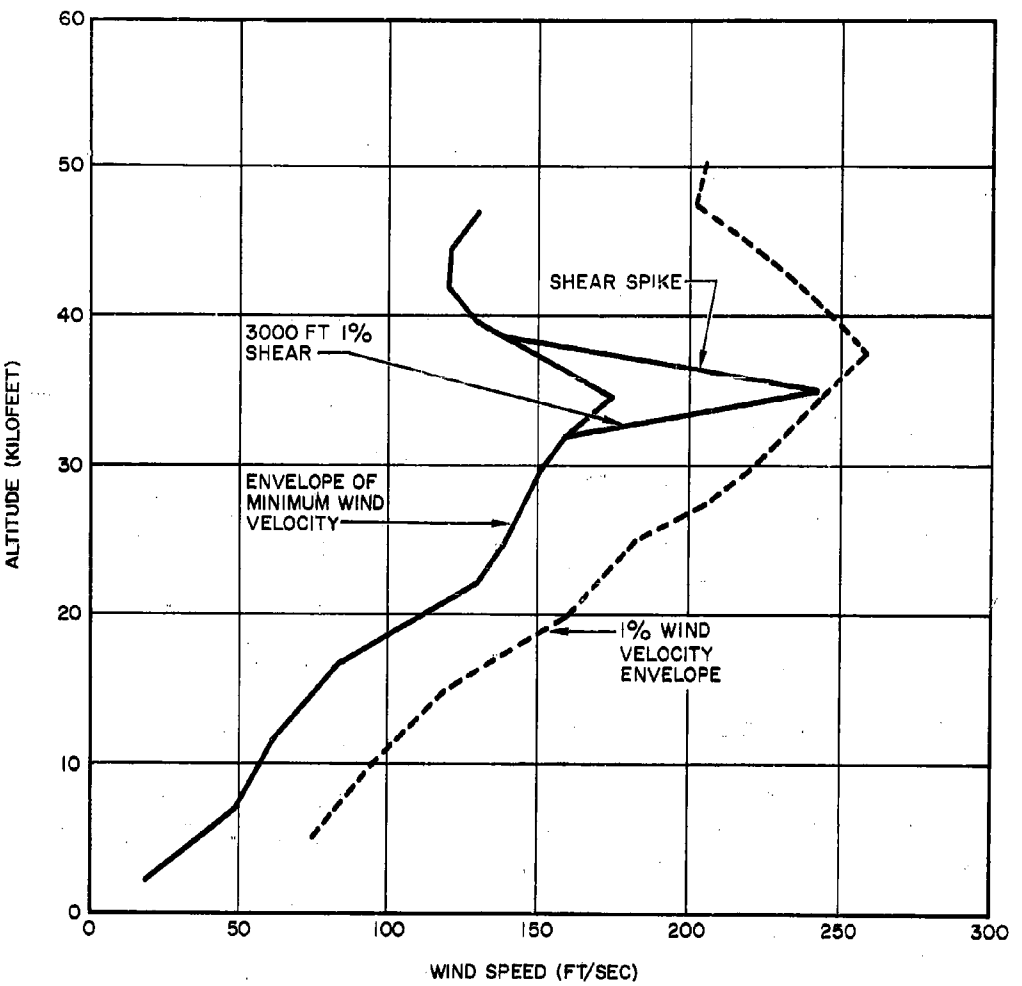


Figure 3. A Synthetic Wind Profile (Reference 5)

The advantage of the above techniques is that once the coefficients are computed, the wind effects statistics can be computed without simulating the missile through a set of random or synthetic profiles. The disadvantages are:(1) the loss of accuracy due to the assumptions that the wind components are normally distributed and that the equations of motion are linear, which allows superposition to be applied; (2) the difficulties involved in checking the answers for reasonableness due to the artificiality of the techniques; and (3) the conceptual difficulties encountered by one not trained in statistics.

5.3 Profile Sample Technique

This technique computes the statistics of the wind effects by simulating the missile flight through an independent random sample of actual profiles. It was first used by Avidyne Research, Inc. (4).

The advantage of this technique is that sufficient information can be obtained from a random sample regarding all wind effects to completely describe their statistics at every altitude with any desired confidence and accuracy. Also, this technique is very closely tied to reality so that answers can be checked and a "feel" for the effects of winds can be obtained.

The chief disadvantages are:

(1) The confidence limits for any given accuracy are a function of the number of profiles used, and the cost of the simulations can become critical when high confidence is required.

(2) A change in missile design or trajectory can change the statistics of the wind effects for the new design and can thus necessitate the repetition of a large number of simulations.

The design method presented in this report will utilize the profile sample technique. Therefore, its use and possible techniques for overcoming the above disadvantages are discussed in detail in the following section.

6.0 USE OF THE PROFILE SAMPLE TECHNIQUE

The proposed methodology uses a set of independent random profiles for computing the effects due to wind profiles. The use of these profiles involves (1) selecting the profile sample, (2) simulating the missile flight through the profiles, (3) reducing the data to a useful form, and (4) checking the effect of any missile design and trajectory changes. These steps will now be discussed individually.

Step 1: Selection of Profile Sample

Wind soundings by means of balloons have been performed at regular intervals (typically, four times a day) at a large number of locations for many years. Avidyne Research, Inc., has obtained from this mass of data an independent random sample of 200 profiles for the winter months at seven locations that are

more or less evenly distributed throughout the United States, (5) and (9). Samples are also being prepared for a number of locations outside the continental United States.

The sample selected for any given missile design would be from the location that would be most representative of the launch site. If the missile was to be launched from several sites, only the most severe of the representative samples is usually considered. However, if it is not clear which sample will produce the severest effects, it may be necessary to use several samples.

The size of the sample is dependent upon the accuracy and confidence limits that are desired. A sample size of 200 was chosen by Avidyne to provide a reasonable compromise between the amount of computation required, and the accuracy and confidence level obtained. A relationship which is used in determining the sample size is

$$n = \left(\frac{\beta}{acc}\right)^2 \left(1 + \frac{a^2}{2}\right)\left(\frac{\mu}{\sigma} + a\right)^2$$

where

n = Size of sample taken from the normal distribution, f(x), with mean, μ, and standard deviation, σ

β = Number of standard deviations required such that the cumulative normal distribution evaluated at β equal (C+1)/2

C = Confidence = Probability that the desired probability of exceedance will fall within the specified accuracy when sample of size n is taken from f(x)

acc = Specified accuracy = Maximum allowable error of $\mu + a\sigma$

a = Number of standard deviations needed to reach desired probability of exceedance.

The above equation assumes that the population distribution is normal and that the ratio of μ/σ is known. As an approximation, the ratio of sample mean to sample standard deviation is used instead of μ/σ for the evaluation of n when μ/σ is not known. If the distribution of the population is not known, the sample size for a given accuracy and confidence level increases greatly.

For an accuracy of 5%, a ratio of μ/σ of 1.7, an α of 2.33 (1% probability of exceedance), and a β of 1.65 (90% confidence), the equation given above gives a sample size, n, of 200. If the form of the probability distribution is unknown, a sample size of around 2500 is needed under the same conditions (10).

Step 2: Simulations

Since the number of simulations can usually be cut only at the sacrifice of confidence limits, the cost of simulations can best be reduced by optimizing the simulation techniques. This optimization could include a trade-off between computer time and simulation accuracy. However, care must be taken since a simplification may have effects on accuracy that vary widely with missile characteristics. Digital simulations are most common, although an analog simulation could prove less expensive if proper function generators could be used, and particularly if several simulations with small differences in missile characteristics were desired for each profile.

Step 3: Data Reduction

Enough data is computed by these simulations to establish probability distributions for the magnitude of all wind effects at every altitude or altitude increment, and in both pitch and yaw. However, the probabilities of interest to a missile designer are the probability of failure or launch delay due to winds aloft - both for any given missile design. How the data from the simulations can be used to compute the probabilities will be discussed later in this report.

Step 4: Recomputing Statistics for Change in Missile Configuration

Changes will often be made to the missile configuration or trajectory during the design process that will appreciably change the wind effect statistics. If these changes occur after the original simulations through the profile sample were completed, the designer is faced with the alternatives of repeating the full number of simulations or of attempting to extrapolate to a new set of statistics by a less costly and probably less accurate technique.

Several possible techniques for avoiding a repetition of simulating through the complete original sample are discussed below. Extreme caution should be exercised in applying any of these techniques to a new situation.

(1) Profile Selection by Assumption,(5) and (9).

The assumption is made that the profiles that gave the largest wind effects for the original vehicle configuration will still give the largest effects for the new configuration.

The new configuration is then simulated through perhaps 25% of the original sample to produce the upper part of the cumulative probability curve.

(2) Profile Selection by Simplified Simulation, (3).

Here the selection of the worst profiles for the new configuration is accomplished by using quasi-steady equations to quickly estimate the wind effect magnitudes from each of the profiles in the original sample. Then, by complete simulation through these worst profiles the interesting upper part of the cumulative probability curve can be constructed.

(3) Sector-Profile Method, (6)

The assumption is made that a given profile in a given compass sector will produce a wind effect magnitude with a probability of exceedance which does not vary for any member of a class of missiles. In other words, the magnitude of the wind effect produced by the profile may change, when the configuration or trajectory of a nominal missile is perturbated, but the statistical meaning of the effect produced will remain the same. This technique was developed for the unsymmetrical missile configuration and includes the solution for a symmetrical missile as a special case.

7.0 DISCUSSION OF PROPOSED DESIGN METHODOLOGY

The terms used in discussions of the methodology were defined in 3.0. However, a discussion of these terms is provided in the following paragraph to further clarify their meanings.

A missile encounters effects due to a number of factors such as wind profiles, profile detail, and buffeting; all of these factors except wind profiles are classified as non-wind profile factors. All factors cause effects (structural loads, control system perturbations, etc.), but when these effects are caused by a wind profile they are termed wind profile effects. Parameters which characterize the effects due to both wind profile and non-wind profile factors are termed effect parameters (bending moment, thrust vector angle). The reason for separate classifications of wind profile effects and non-wind profile effects is that they are handled differently in the methodology. Wind profile effects are computed from simulations through a random set of wind profiles. The statistics of these effects are never explicitly described. Non-wind profile effects are computed by considering the amplitudes of each effect parameter resulting from all non-wind profile factors in an interval of flight time. These effects are described by probability distributions of the maximum amplitude of each effect parameter in each time interval.

The heart of the design methodology being presented are the methods for the computation of probability of failure or launch delay for a given design. (These methods can also be used for studying the capabilities of an existing

missile). The methods are unique in that they consider random factors other than wind profiles and profile detail, they divide the flight into intervals of time, and they consider a probability distribution of only the maximum values of non-wind profile effects in a time interval.

The method for computing the probability of failure uses results of flight simulations through a set of random wind profiles and also probability distributions for non-wind profile effects in each interval. The probability of failure is computed first for each random profile in each interval of flight time (see Figure 4), then for each profile during an entire flight, and finally for all profiles in a random sample.

The method of computing the probability of launch delay first determines the launch allowable values for wind profile effects. This is done by subtracting an allowance for non-wind profile effects from the design allowable in each time interval (see Figure 5). The method then finds the probability that a random wind profile will cause a launch allowable to be exceeded.

Since the probabilities of interest can only be found for a given design, the process of designing to the criteria of 4.0 is necessarily iterative. The design may have to be changed several times before one with the desired probability value, or with an optimized probability, can be found. This is a difficult process requiring use of large computer installations and a good deal of engineering judgment. However, it is often worthwhile because of the importance of obtaining a good design.

The advantages of the proposed methodology are (1) that computational steps are closely tied to reality, and (2) that the amount of computation can be varied according to the degree of accuracy desired in the design. The disadvantage is that two major simplifying assumptions regarding non-wind profile effects have to be made to keep the necessary computations from becoming more complex and costly than is justified for the amount of basic data now available.

The two main assumptions used in the methodology are stated and discussed below. Assumption 1 can be modified under special conditions to be less inclusive. A clarification of this statement will be included in the discussion of Assumption 1.

Assumption 1: Unity Correlation Exists Between All the Probability Distributions of all the Effect Parameter Magnitudes Resulting from Non-Wind Profile Factors.

The assumption means that, when non-wind profile effects are considered, there is unity correlation between two different effect parameters in the same time interval, between two different effect parameters in the different time intervals, and between the same effect parameter in two different time intervals.

This assumption is necessary because the methodology usually divides the period of flight in which the probability of failure due to winds is appreciable into a number of time intervals (see 8.0), and because the

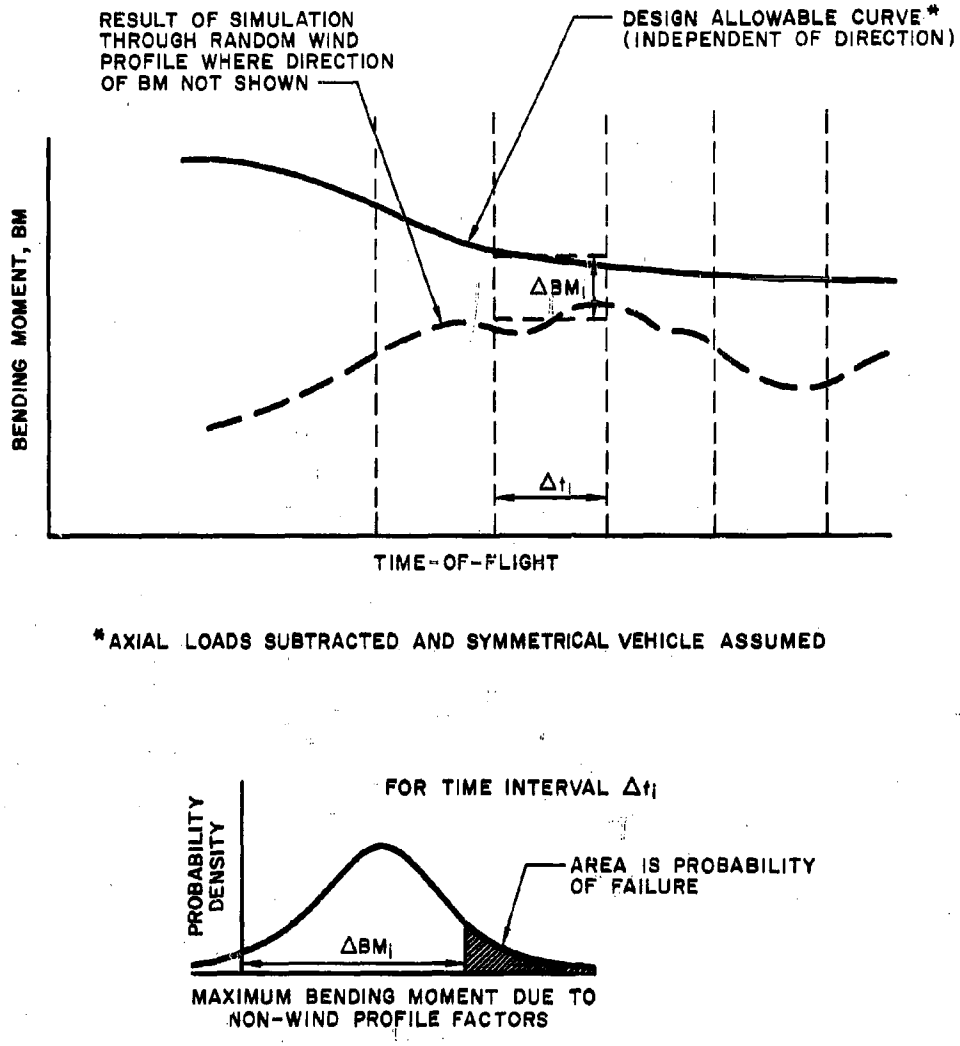
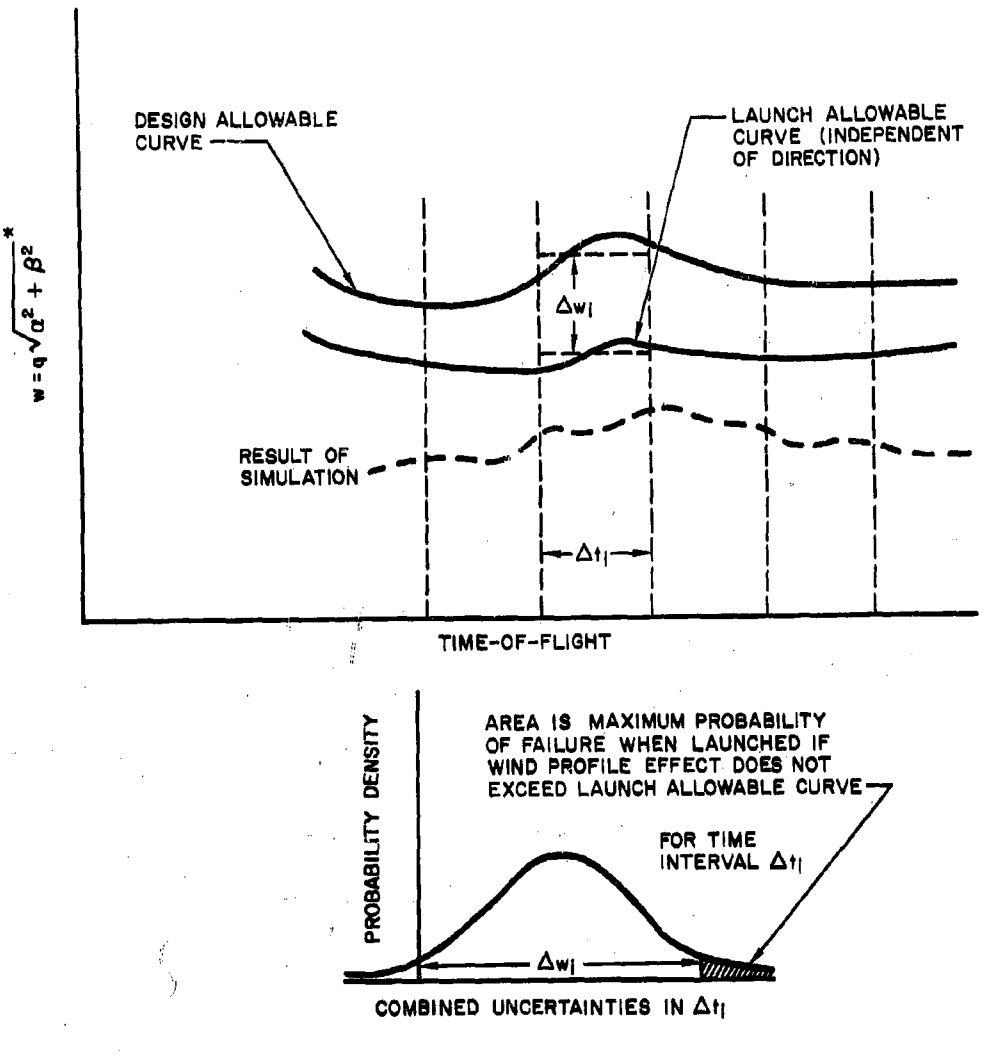


Figure 4. Graphical Solution of Probability of Structural Failure



* q IS DYNAMIC PRESSURE
α IS ANGLE OF ATTACK
β IS SIDESLIP ANGLE

Figure 5. Graphical Solution of Launch Allowable Curve

methodology considers the possible failure of the missile to be due to any effect parameter exceeding its design allowable. In the methodology, the wind profile effects versus flight time are computed for each profile from the random wind sample being used; and, since the probability of failure is first computed for each profile, the wind profile effects are used as deterministic values. However the effect parameters for non-wind profile effects are probable, and correlations between them in one time interval or two different time intervals must be used in the methodology. It is known that these correlations are high, but in the present state-of-the-art they cannot be computed. (Since the dominant non-wind profile effects are usually due to profile detail, additional statistical data regarding profile detail is particularly needed.)

The assumption of unity correlation simplifies computations because it means that the probability of failure for one profile is due to the highest probability of any effect parameter in any time interval exceeding its design allowable. Without such an assumption, multivariate distributions must be utilized; and published tables describing such distributions are available only for the cases of two or three variables. Thus, only in the special case of an unusually small number of time intervals and/or critical effect parameters, could this methodology be used in the practical case without the assumption of unity correlation.* Although approximations could be made in any problem to decrease the number of variables, the inaccuracies could be as damaging as the unity correlation assumption itself. Therefore, the methodology described is only for the condition that Assumption 1 is being used.

The most undesirable feature of the assumption is that it is unconservative -- it reduces the computed probability of failure. Therefore, the effect of the assumption must be examined carefully for any specific probability computation. The errors which will result from the assumption are proportional to the number of time intervals used and to the number of effect parameters that have an appreciable probability of exceeding their design allowable. Also, the absolute errors decrease as the magnitude of the non-wind profile effects decrease, for it must be remembered that this assumption does not apply to the consideration of the usually dominant wind profile effect magnitudes. Finally, although a number of time intervals are considered, often the probability of failure in one time interval is much larger than in the other time intervals for most of the profile simulations. Such a situation also decreases the errors due to the assumption.

Assumption 2: The Magnitudes of an Effect Parameter Caused by all Non-Wind Profile Factors can be combined Algebraically with Each Other and with the Magnitude of the Same Parameter Caused by a Wind Profile.

This assumption means that if a wind profile produced an effect parameter such as a bending moment in one plane of a missile (e.g., pitch plane), then the bending moment due to all non-wind profile factors (such as profile detail, buffeting, etc.) also occurs in that same plane. Thus, the effects can be combined by simple addition or subtraction.

*An exception results from a new technique described in (11).

This is a common assumption in wind studies. It can be justified on the basis that (1) it is conservative, (2) the data for more exact computations is not available, (3) the consideration of random directions would probably be prohibitively difficult, and (4) the most important non-wind profile factor, profile detail, is believed to exist mainly in the plane of the wind profile. In practice, what a designer must do is to estimate equivalent non-wind profile effect magnitudes in the compass sectors containing the dangerously high wind profile velocities.

8.0 COMPUTING THE STATISTICS OF NON-WIND PROFILE EFFECTS

The methodology of this report requires the determination of the probability distribution of the maximum magnitude of each critical effect parameter resulting from all non-wind profile factors for each interval of flight time in which winds are an important factor. (Such a distribution is shown in Figure 4.) To do this, one must first choose the time intervals. Then, for each interval, one must compute the probability distribution of each effect parameter for each factor. Finally, a combined distribution of each effect parameter for all factors is determined in each interval. A detailed discussion of these three steps follows.

Step 1: Selection of Time Intervals

The period of flight of interest is only that in which there is an appreciable probability of failure due to winds. It is assumed that the probability of failure at any other time-of-flight is zero. The period is then divided into time intervals such that the probability of failure during each interval is approximately constant with time.* The shorter the interval the more constant is the probability, but it is desirable to use long intervals in order to simplify computations and reduce inaccuracies due to Assumption 1 in 7.0. The choice of the intervals is therefore often a compromise, and it is always an estimate. One must consider variations with time of the design allowables, the non-wind profile effects, and the wind profile effects. An example of a large change in non-wind profile effects with time would be the advent of buffeting in the region of Mach 1. Wind profile effects will also change around Mach 1 due to changes in aerodynamic coefficients. Typically, the period of flight from 25K to 40K feet altitude is divided into time intervals of 10 seconds or more.

Step 2: Computation of Statistics for Each Effect Parameter Magnitude Resulting from Each Non-Wind Profile Factor

For each effect parameter of interest (e.g., bending moment), a probability distribution of its maximum amplitude must be found for each non-wind profile factor (e.g., profile detail) in each time interval. These distributions will be approximated as normal probability distributions, and the necessary data for the approximations are obtained from past flight tests, simulations, etc. Normal distributions are approximated in order to simplify the combination of distributions that will be discussed in Step 3.

*Each interval is therefore representative of one phase of flight.

Since some correlation exists between profile detail (the dominant non-wind profile factor) and wind velocity, a different probability distribution should be used for each level of wind profile velocity. In other words, conditional probability distributions could be approximated for each condition of wind profile - or even for each condition of wind profile effects. The data for such conditional probabilities does not now exist. An alternative to the use of a number of conditional distributions is to use a conditional distribution for a high wind profile velocity or profile effect, since only under such conditions is there usually an appreciable probability of failure. This alternative will cause a conservatively large probability of failure to be computed.

The computations for one effect parameter (structural bending moment) due to one non-wind profile factor (profile detail) in one time interval are discussed in the following paragraph to clarify the procedures necessary for this step.

Assume that wind velocity and wavelength statistics for profile detail are available for all altitudes of interest. (These statistics should be for the a priori condition of a high wind profile velocity.) A fixed time simulation (the velocity vector, altitude and missile parameters are fixed) is then performed at a representative time-of-flight in the time interval with the missile being excited by the power spectrum of the aerodynamic forces due to profile detail. This simulation is continued for time equal to a large multiple of the time interval, and the bending moment at the critical missile station (the location with the highest probability of having its design allowable exceeded) is recorded. This recording can then be divided into increments equal to the time interval under consideration, and the maximum positive (negative) bending moment during each increment is measured. It is important to consider only all positive or all negative moments, for in flight only one will add to the wind profile effect. Consideration of a maximum absolute magnitude could produce distorted results. Either positive or negative can normally be used since the effects are usually symmetrical. An illustration of results from such a simulation is shown in Figure 6. Note that the period of the dominant frequency of the bending moment (usually a result of the first flexible body mode) is considerably shorter than the time interval. This makes the concept of maximum magnitude in the interval more meaningful.

A normal probability density can be approximated for the measured maximum bending moment. This distribution will have a non-zero mean, as is shown in Figure 6.

Step 3: Computation of a Combined Distribution in Each Interval for Each Effect Parameter Magnitude Resulting from All Non-Wind Profile Factors.

A probability distribution from Step 2 of the maximum amplitude of one effect parameter for one non-wind profile factor in a time

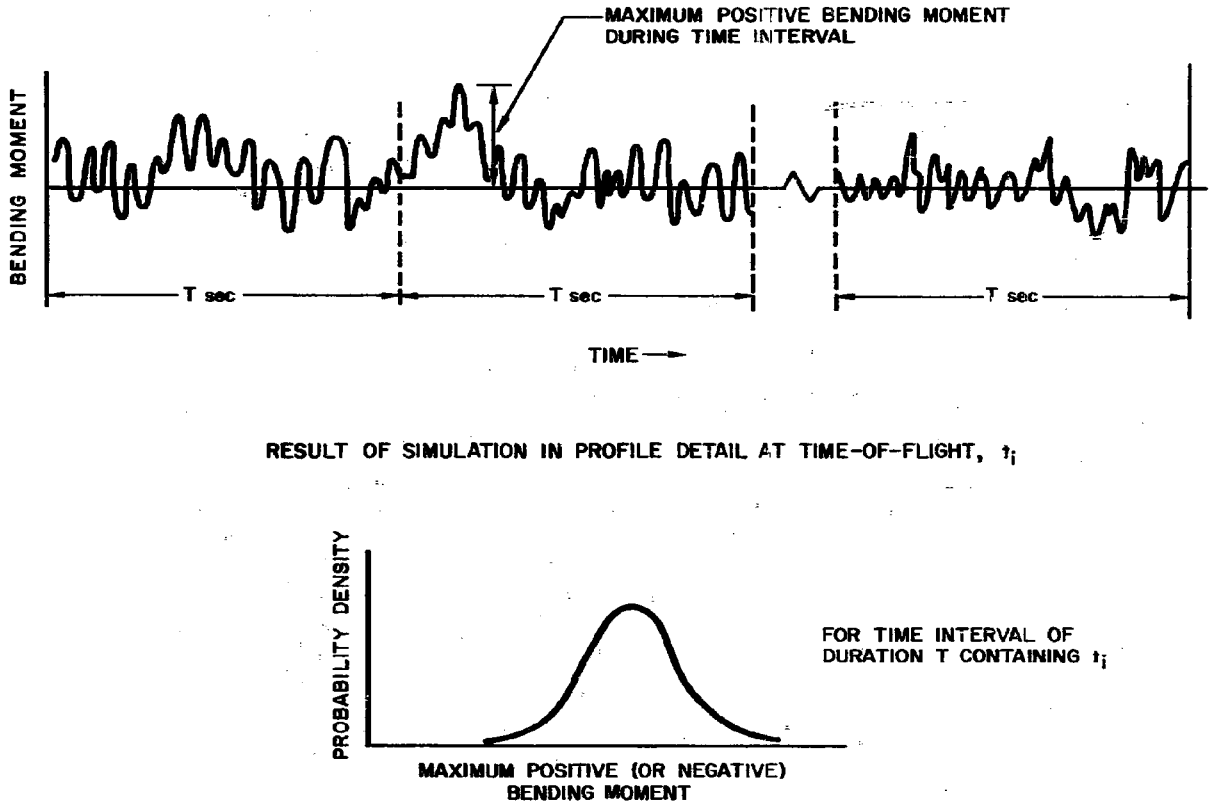


Figure 6. Obtaining Statistics of Structural Loads Caused by Profile Detail

interval must be combined with the distributions of the same parameter caused by all other non-wind profile factors in that interval. Since all are approximated as normal distributions, and since all are assumed to act in the same direction (Assumption 2 in 7.0), this can be accomplished by the well-known formulas given below. This must be done for each effect parameter in each time interval.

$$\text{Combined Mean} = \sum_{j=1}^{N} \mu_j$$

$$\text{Combined Variance} = \sum_{j=1}^{N} \sigma_j^2 + 2r_{jk} \sum_{j=1}^{N} \sum_{k=1}^{N} \sigma_j \sigma_k$$

where μ_j = mean of jth probability distribution

σ_j^2 = variance of jth probability distribution

r_{jk} = correlation coefficient

$j \neq k$

N = number of combined variables.

The factors listed in Section 3.2 are essentially independent, so that the correlation coefficient, r_{jk}, is normally zero; however, the methodology does not require independence.

9.0 DESIGNING FOR PROBABILITY OF FAILURE

As previously defined, the probability of failure is the probability that the design allowable of any effect parameter (bending moment) will be exceeded at some time during flight. The methodology for designing to a given probability of failure, or to an optimum probability of failure, is outlined in the steps listed below. If it is desired to check an existing missile, the last step can be deleted. In the methodology, the probability of failure will be computed first for each random profile in each time interval for all effect parameters, then for each profile during an entire flight, and finally for all profiles in the random sample.

Step 1: Computation of Design Allowables

A specified design allowable curve (defined in Section 3.3) versus time (during the period in which winds can endanger the flight) is needed for each effect parameter that may have an appreciable probability of exceeding its allowable. (Several allowable curves will be needed for each effect parameter for an unsymmetrical vehicle as is shown in Appendix I.) Parameters that could be critical are discussed in Section 3.0. Usually only one or two effect

parameters need be considered for any one vehicle. The computation of design allowables is a normal part of a missile design and will not be discussed here.

Step 2: Selection of The Sample of Random Wind Profiles

An independent random sample of profiles which are obtained from balloon soundings must be chosen. The location of the balloom launching site and the size of the sample must be carefully considered as described in Section 6.0.

Step 3: Simulations Through Sample Profiles

The flight of a nominal missile of a given design must be simulated through each of the profiles. The time history of each effect parameter of interest from each simulation can be plotted on graphs showing the design allowable curves. Such a graph is shown for a symmetrical structural load in Figure 4. See Appendix I and Figure 7 for consideration of an unsymmetrical design allowable curve.

Step 4: Selection of Time Intervals

Any period of flight in which the effects of winds are critical may have to be divided into time intervals. The choice of time intervals was discussed in Section 8.0.

Step 5: Computation of Probability Distributions for Non-Wind Profile Effects

The combined probability distributions for the maximum amplitude of each effect parameter of interest resulting from all factors except wind profiles must be computed for each time interval. Such computations were discussed in Section 8.0.

Step 6: Determination of Average Difference Between Wind Profile Effect Magnitudes and Design Allowables in Each Time Interval for Each Profile

If the magnitude of an effect parameter resulting from only the wind profile exceeds a design allowable value in any time interval, it is clear that the probability of failure for that profile is unity. However, even if a wind profile alone does not cause failure, there is always some probability that the additive non-wind profile effects can cause failure. To compute that probability for each profile, the difference between the wind profile effect magnitude and the design allowable must be computed in each time interval for each effect parameter of interest. Since both may be changing during the interval, some discretion must be exercised in this computation. The difference between the averages is usually the best approximation, as is used in Figure 4.

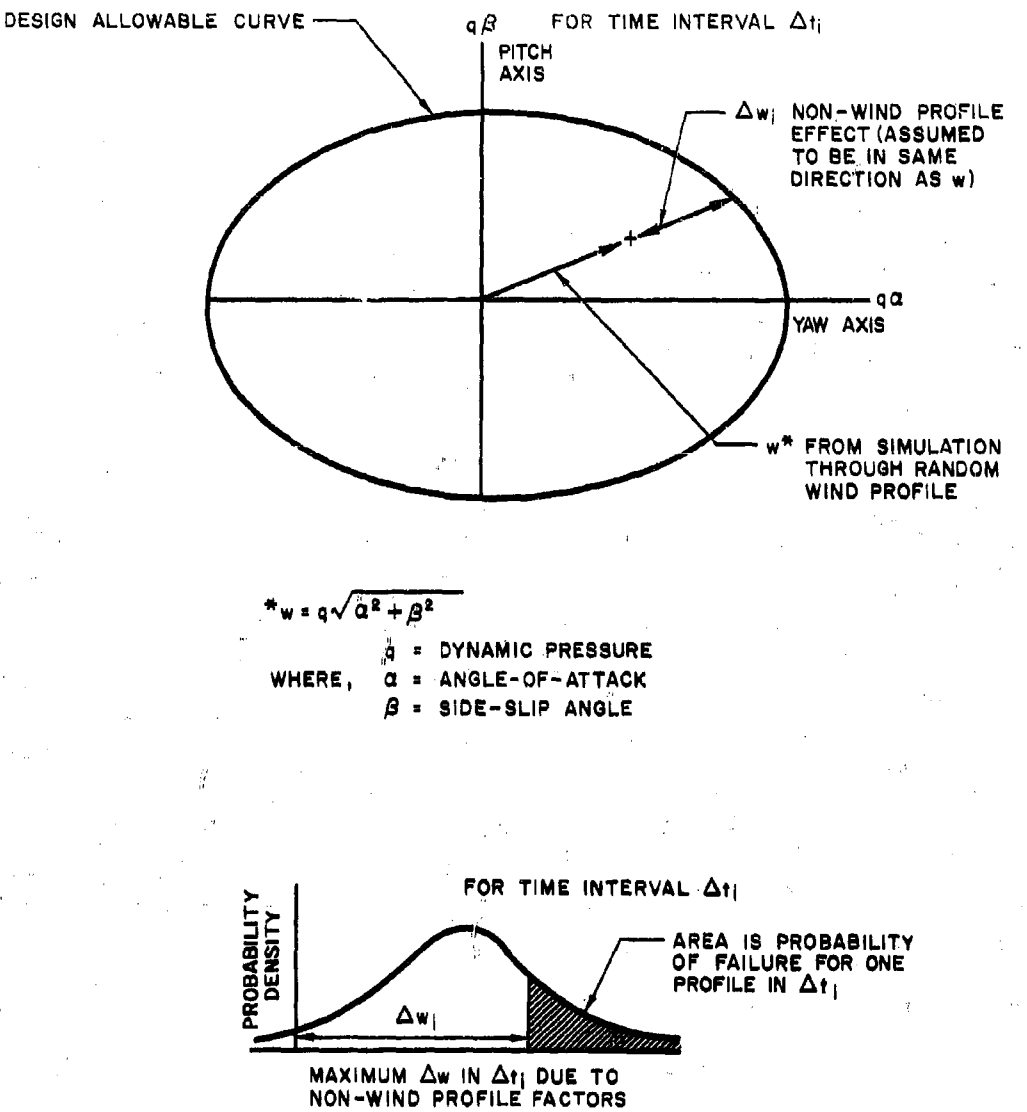


Figure 7. Graphical Solution of Probability of Structural Failure for Unsymmetrical Vehicle

Step 7: Computation of the Probability of Failure in Each Time Interval for Each Profile

The probability of failure for each profile in each time interval is determined by first finding the probability that the magnitude of each effect parameter resulting from non-wind profile factors will exceed its difference value found in Step 6. These probabilities can be obtained graphically by integrating the area under the part of the effect parameter probability distribution curve that exceeds the difference (as shown in Figure 4). They can also be obtained from tables of the cumulative normal distribution. This step utilizes Assumption 2 in 7.0. Using Assumption 1 in 7.0, the probability of failure in each time interval for each profile is simply the highest probability of any effect parameter exceeding its differential value due to all non-wind profile factors.

Step 8: Computation of the Probability of Failure for Each Profile

If Assumption 1 in 7.0 is used, then the probability of failure for each profile for any effect parameter during an entire flight is simply the highest probability of failure in any time interval for any effect parameter. (If Assumption 1 is not used so that correlations are arbitrary, then multivariate distributions must be utilized as described in the discussion under Assumption 1 in 7.0.)

Step 9: Computation of the Probability of Failure for the Profile Sample

The probability of failure associated with each profile in the selected random sample was found in Step 8. The probability of failure for the entire sample is the sum of the probabilities for the individual profiles divided by the number of profiles in the sample. This is the actual probability of failure for the given design within the accuracy and confidence limits determined by the size of the sample.

Step 10: Iteration of the Design to Meet the Design Criteria

If one is designing to a given probability of failure, and if Step 9 shows that the original design does not meet the criteria, then the design must be changed. If the change is of a type that causes the wind effects to change, then they must be recomputed as discussed in Step 4 of Section 6.0. In any case, the new design allowables must be found, and a new probability of failure computed. Still further iterations could be necessary, although rapid convergence to a reasonable accuracy of design would be expected.

If one is designing to an optimum probability of failure, some utility function (criterion used in the optimization of a system) must be established to allow a trade-off between probability of failure and some other parameters (such as weight, cost in dollars, schedule slippage, etc.). An iterative design procedure would then be followed.

10.0 COMPUTING LAUNCH ALLOWABLES

Many large boosters carrying space vehicles are not designed to withstand extreme wind conditions which have an appreciable probability of occurrence. Since it is usually preferable to delay the flight under such conditions rather than risk a failure, prelaunch wind checks are performed. Based on these checks (normally in the form of balloon soundings), predictions are made of the magnitude of wind profile effects that will be encountered in flight. If the predicted values exceed a launch allowable (defined in 3.5), then the launch is delayed until the predicted values decrease. Such a procedure is described in Appendix II. The launch allowables are so selected that, when a flight is permitted on the basis of the allowables, the probability of failure due to winds in that flight is less than some assigned value (usually a negligible value).

Launch allowable values are assigned only to wind profile effects, since only they can be predicted with any reasonable accuracy. Thus, statistics of non-wind profile effects must be precomputed, and an allowance for these effects must be subtracted from the design allowables in each time interval. The computation of these non-wind profile effects is discussed in 8.0. However, in order to arrive at the launch allowables, a further allowance must be made for uncertainties in the accuracy of the predicted wind profile effects. These uncertainties are a result of (1) possible inaccuracies in the prediction process, and (2) possible variations in the wind characteristics between the time of the last balloon sounding and the actual launch time.

Inaccuracies in the prediction process are dependent on the techniques used. If the simulations are made through wind profiles measured before launch, errors can occur due to inaccurate data or approximations in the simulation. If only precomputed regression curves (one such curve is shown in Figure 8) are used for prediction, even larger errors could be encountered. Variations in wind characteristics between balloon soundings and launch are dependent on prior wind conditions and upon the length of time involved. Naturally, every effort is made to keep the time between sounding and launch to a minimum in order to minimize these variations. Normal probability distributions are approximated for the effects of these new uncertainties as was done with the non-wind profile effects, and independence is assumed. Thus their distributions can be easily combined with those of the non-wind profile effects.

The allowances for prediction uncertainties to be subtracted from the design allowables, are found from such a combined probability distribution and reflect all of the flight uncertainties except the wind profile. This probability distribution is computed for each effect parameter in each time interval. A magnitude for each effect parameter in each interval is then chosen with a probability of exceedance equal to the probability of failure that is allowed when the decision to launch is made. These allowances are then subtracted from the design allowables to obtain the launch allowables. An illustration of this procedure is shown in Figure 5.

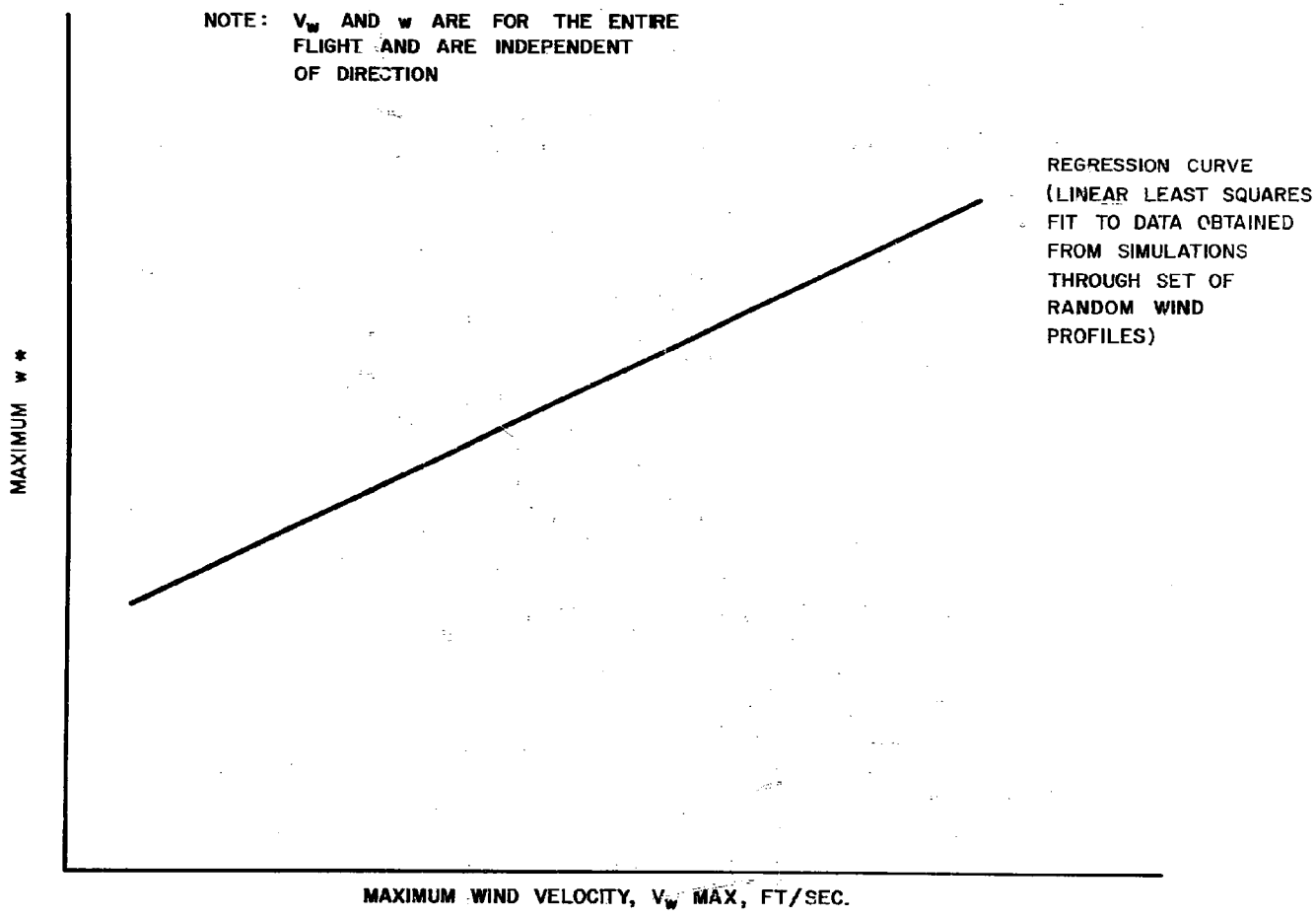


Figure 8. Regression of Structural Load Versus Wind Velocity

If it is assumed that unity correlation exists between all the combined probability distributions of uncertainties (similar to Assumption 1 of 7), then the probability of failure for an allowed flight cannot be more than the probability of failure in any one time interval.

11.0 DESIGNING FOR PROBABILITY OF LAUNCH DELAY

The computation of launch allowables is discussed in 10.0. Since a launch will be delayed whenever a wind profile effect predicted by some procedure before the launch exceeds a launch allowable, it is necessary to compute the probability of such an event for a given missile design in order to perform a design for some probability of launch delay. (one procedure for predicting wind profile effects is given in Appendix II.) The computation of a probability of launch delay for a given design as performed in the proposed methodology usually makes use of the assumption discussed below. The discussion includes a short description of a methodology to be used if the assumption is not utilized.

Assumption: The Wind Profile Effect Statistics Obtained by Simulations Through a Set of Random Wind Profiles Will Be Representative of a Random Set of Predicted Wind Profile Effects.

The assumption means that the same statistical parameters for the magnitude of an effect parameter, such as bending moment, would be obtained from a random sample of prelaunch predictions of bending moment as from simulations through a random sample of profiles.

This assumption is valid when there is no error in the wind effect predictions, but this will never be completely true. No bias will usually exist in the predictions, so the mean value of a predicted effect parameter is the same as the mean obtained from simulations using random profiles. However, there will be errors in the predicted values that can cause the standard deviation of a predicted effect parameter to be greater than the standard deviation obtained by using the random profiles. This larger standard deviation means that the predicted wind effects will exceed a launch allowable a larger percentage of the time, and the probability of launch delay will increase. Therefore, the probability computed on the basis of this assumption is unconservative.

Not only is the assumption unconservative, but it is inconsistent with the allowance for a prediction error that was used when computing the launch allowable curve in 10.0. However, the allowance for prediction errors should remain even if this assumption is used, for minimizing flight failure is certainly more important than an inaccuracy in a computed launch delay probability.

The assumption is used for many of the reasons stated for previous assumptions, namely: it simplifies the computations,

data for an accurate computation is not available, and, the error encountered in its use should not be appreciable in mose cases.

The magnitude of error that results from the assumption decreases as the accuracy of the prediction process increases. If it is planned to use a complex procedure such as described in Appendix II, then the assumption should be reasonable. However, the planned use of a relatively crude procedure could make it necessary to consider the prediction error. A method of doing this is described in the following paragraph.

If the above assumption cannot be used, then the probability of launch delay can be computed in the same manner as was the probability of failure in 9.0. In this case, however, an estimate of the probability distribution for each effect parameter due to prediction errors must be added to the result of each profile simulation in each time interval. The probability of exceeding the launch allowable can then be computed in each interval for each profile, then for each profile, and finally for the entire profile sample. These computations can be simplified by the use of an assumption similar to Assumption 1 in 7.0 - that is, unity correlation can be assumed between all distributions of the effect parameter magnitudes resulting from prediction errors.

The methodology for designing to a probability of launch delay or optimizing the probability is outlined in the following steps using the foregoing assumption.

Step 1: Selection of Time Intervals

This is done as described in Section 8.

Step 2: Computation of Launch Allowables

See Section 10.

Step 3: Selection of Sample of Random Wind Profiles

See Section 6.

Step 4: Simulations Through Sample Profiles

The flight of a nominal missile of a given design through each of the profiles must be simulated.

Step 5: Comparison of the Results of the Simulations with the Launch Allowable Curves

The time history of each effect parameter of interest from each simulation can be effectively plotted on a graph showing its launch

allowable curve. This is illustrated for a symmetrical vehicle in Figure 5. See Appendix I and Figure 9 for consideration of an unsymmetrical vehicle. The graphs will show when an effect parameter exceeds its launch allowable curve.

Step 6: Computation of the Probability of Launch Delay

The fraction of the random profiles that cause a launch allowable for any effect parameter to be exceeded is the probability of launch delay with an accuracy and confidence proportional to the size of the sample of profiles.

This computation is meaningful when the following statements are true:

(1) A launch delay decision is always based solely on the criteria that a predicted effect exceeds its launch allowable.

(2) The decision to delay a launch is always made at one scheduled time in the countdown.

This statement must be assumed to be true in order to avoid consideration of the increased probability of launch delay that would be encountered when wind profile effect predictions and launch decisions are made several days before launch. (See Appendix II.) It is normally true because, in an actual situation, the launch delay decision is usually postponed to just before launch in the hope that winds may improve, and because persistence of winds makes the probability of delay relatively insensitive to decision time over a period of several hours.

Step 7: Iteration of the Design to Meet the Design Criteria

If one is designing to a given probability of launch delay, and if Step 6 shows that the original design does not meet that criteria, then the design must be changed. If the change is of a type that causes the wind effects to change, then they must be recomputed, as discussed in Section 6. In any case, the new launch allowables must be found and a new probability of launch computed. Still further iterations could be necessary, although rapid convergence is expected.

If one is designing to an optimum probability of launch delay, some utility functions must be established to allow a trade-off between probability of delay and some other parameter (such as weight, cost in dollars, schedule slippage, etc.). An iterative design procedure would then be followed.

12.0 CONCLUDING REMARKS

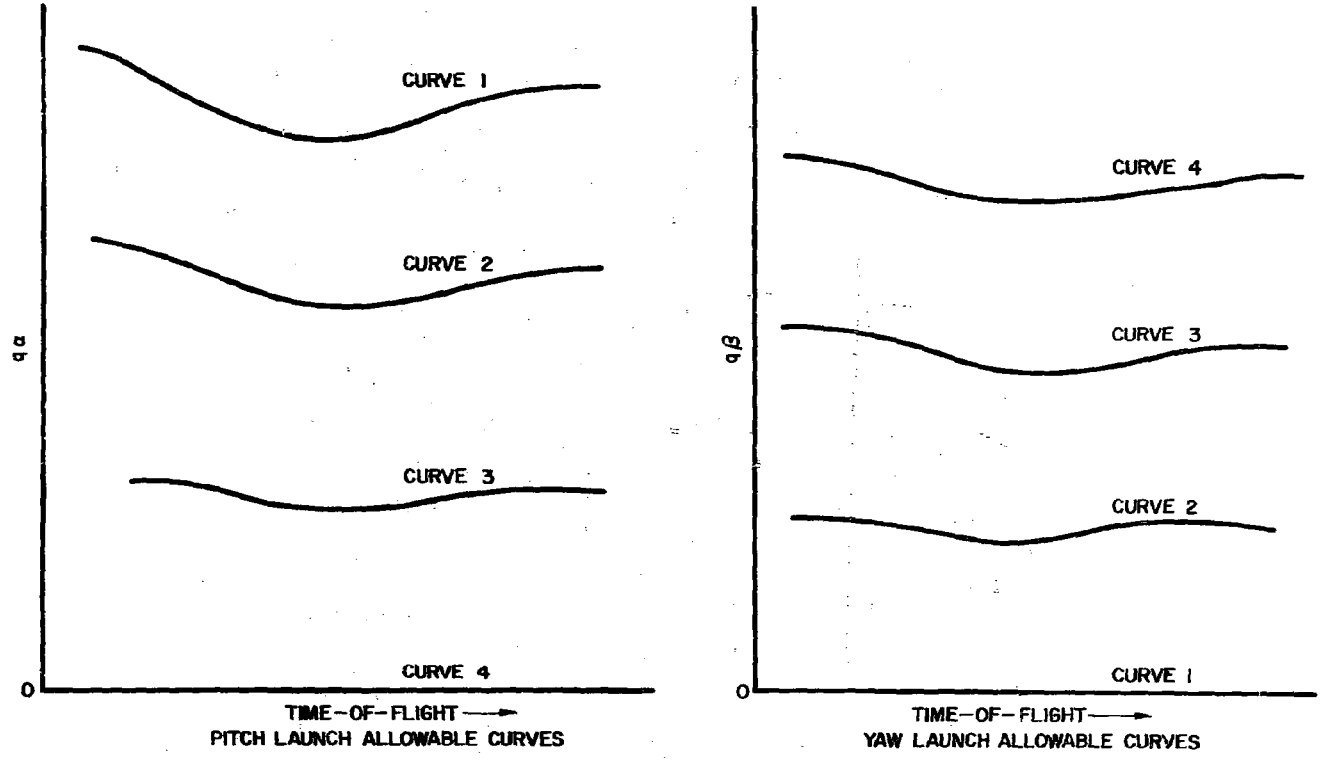
The presentation of this methodology has attempted to consider the general case--that is, consideration of all possible vehicle configurations and a range of requirements on the desired accuracy of the results. This

TO USE CURVES: THE STRUCTURAL LAUNCH ALLOWABLE CURVE IS EXCEEDED IF, AT ANY GIVEN ALTITUDE:

(1) CURVE 1 IN PITCH OR CURVE 4 IN YAW IS EXCEEDED,

OR

(2) IF $q\beta$ IS ON A HIGHER NUMBERED CURVE THAN $q\alpha$. (MORE CURVES CAN BE OBTAINED BY INTERPOLATION FOR MORE ACCURACY.)



NOTE: FOR A SYMMETRICAL VEHICLE, CURVE 1 EQUALS CURVE 4, ETC.

DEFINITIONS: q IS DYNAMIC PRESSURE
α IS ANGLE-OF-ATTACK
β IS SIDESLIP ANGLE

Figure 9. Structural Launch Allowable Curve for Unsymmetrical Missile

makes the procedures appear more complex and burdensome than they need be for a specific situation. Actually, the mechanics of performing some of the steps can be simplified by tailoring them to one's needs and computing facilities.

Studies are being conducted at Aerospace in an effort to develop a design methodology that utilizes the concepts presented here, but uses reduced statistical data from the simulations through the random profiles instead of dealing with one profile at a time. Apparent advantages are increased mechanization of computation and greater flexibility in the assumption of correlation coefficients. The main disadvantage appears to be a loss of physical reality. Such a loss makes it more difficult to check answers for reasonableness and to apply engineering judgment.

It is of interest to examine the type of additional meteorological data that would improve the accuracy of the type of methodology presented here. The most valuable data would describe profile detail, and should include the statistics of wavelength and magnitude as well as the correlations between altitudes with wind profile characteristics.

In connection with the gathering of data on profile details, there is some question as to how wind data should be divided to separately describe wind profiles and profile detail. There are arguments for continuing to have the statistics for profile detail include all wind characteristics that have shorter wavelengths (wavelength is in feet of altitude) than the wavelengths detected by present balloon sounding techniques. The alternative is to use balloon sounding techniques that detect wind changes over smaller altitude layers, and to include this data in the statistics that describe the wind profiles. The separately defined statistics for profile detail will then include wind characteristics with a narrower wavelength spectrum. The arguments for maintaining the present wavelength division of wind statistics are: (1) a great mass of balloon sounding data already exists, and (2) there are differences in the effects and persistence characteristics of the smoothed profile and the profile detail. The effects differences are largely due to the fact that almost all the dynamic effects come from the profile detail, as contrasted with the quasi-steady state effects of a smoothed profile. This usually demands a division of analysis techniques. The difference in persistence characteristics (these differences are yet to be determined) would be of importance when attempting to predict wind effects to be encountered in a flight. Here again, a division of techniques, one for the smoothed profile and one for detail, may have to be used for the prediction process.

REFERENCES

1. Trembath, N. W., "Missile Control System Design Using Wind Statistics" Proceedings of the National Symposium on Winds for Aerospace Vehicle Design, Bedford, Massachusetts, March 1962.

2. Bieber, R. E., "Prediction of Wind-Induced Missile Structural Loads" Proceedings of the National Symposium on Winds for Aerospace Vehicle Design, Bedford, Massachusetts, March 1962.

3. Clingan, B. E., "A Rapid Method for Determining Wind Shear Design Loads Based on a Set of Measured Profiles" Proceedings of the National Symposium on Winds for Aerospace Vehicle Design, Bedford, Massachusetts, March 1962.

4. Hobbs, H. P., Criscione, E. S., Mazzola, L., and Frassionelli, G., "Development of Interim Wind, Wind Shear, and Gust Design Criteria for Vertically Rising Vehicles," WADC Technical 59-504, July 1959.

5. Mazzola, L., "Design Criteria for Wind Induced Flight Loads on Large Boosted Vehicles" Proceedings of American Rocket Society Conference on Launch Vehicles: Structures and Materials, Phoenix, Arizona, 3-5 April 1962.

6. Schwartz, M. D., "A Method for Computing the Probability of Failure Due to Winds for a Class of Missiles," Aerospace Corporation 1951.2-211, August 1962.

7. Press, H. and Steiner, R., "An Approach to the Problem of Estimating Severe and Repeated Gust Loads for Missile Operations," NACA TN-4332, September 1958.

8. Van Der Mass, C. J., "Wind Shear Response for Missile System Comparative Study and Design Procedure," Lockheed Missiles and Space Co., A003108, 16 January 1962.

9. Mazzola, L. L., Hobbs, H. P., and Criscione, E.S., "Wind, Wind Shear, and Gust Design Criteria for Vertically Rising Vehicle as Computed on the Basis of Montgomery, Ala., Wind Data," WADD TR 6199, April 1961.

10. Schwartz, M. D., "Confidence Intervals using Parametric Estimators and Non-Parametric Techniques," Aerospace Corporation, 1951.2-212, August 1962.

11. Schwartz, M. D., "Evaluation of the Normal Distribution with Many Variables for the Case of Low Probability of Failure," Aerospace Corporation, 1951.2-217, August 1962.

APPENDIX I

CONSIDERATION OF AN UNSYMMETRICAL VEHICLE

The design allowable bending moment curves (defined in Section 3.3) shown in Figure 5 and 6 of this report were for a symmetrical vehicle - that is, the allowable bending moment at the applicable missile station is the same in all directions about the longitudinal axis. Most missiles do tend to be symmetrical in regards to allowable bending moment. However, there is usually enough aerodynamic dissymmetry so that the allowable dynamic pressure, q, times the angle between the body axis and the relative wind velocity, $(\alpha^2 + \beta 2)1/2$, varies with direction (see Figure 7). The importance

of this arises from the fact that the structural load computed in a simulation through a wind profile, whether in a random wind study or as part of a prelaunch wind check, is usually in terms of $q\sqrt{a^2 + \beta^2}$ because it requires additional computation to obtain bending moment at the critical station. Thus design allowable curves and launch allowable curves are often used in terms of $q\sqrt{a^2 + \beta^2}$ and, when they are unsymmetrical, special graphical techniques are required for the methodologies presented in the main body of the report. Handling of the dissymmetry in computing the probability of failure and in performing a prelaunch wind check are discussed here.

The computation of the probability of failure associated with one random profile in one time interval, for an unsymmetric design allowable curve is illustrated in Figure 7. Note that the non-wind profile effects are assumed to add in the direction of the wind profile effect (Assumption No. 2, Section 7). A representative design allowable curve as shown in Figure 7 must be used for each time interval to permit the computation of a total probability of failure.

An assymmetrical launch allowable curve (defined in Section 3.5) is often encountered when performing a prelaunch wind check and making a decision as to whether a launch should be allowed in the existing wind environment. (Such a procedure is described in Appendix II). This procedure often requires that a simulation be made of the vehicle flight through a wind sounding obtained by a balloon sounding made a few hours before scheduled launch time, and that a time history of $q\sqrt{a^2 + \beta^2}$ be recorded from the simulation. The history of $q\sqrt{a^2 + \beta^2}$ must then be compared with the precomputed launch allowable values in order to make a launch decision. If the vehicle is aerodynamically unsymmetrical, then the comparison can be made on a special graph of the type shown in Figure 9. That graph essentially breaks an allowable curve similar in shape to the one shown in Figure 7 into allowable $q\alpha$ and $q\beta$ components for a number of directions. It does that at a number of times of flight so that continuous allowable curves versus time can be drawn. With the simulation history plotted on Figure 9, it is possible to see if it exceeds the allowables at any time and, if not, to see at what time-of-flight it comes closest to the allowable curves. Figure 9 includes directions for interpreting the curves.

APPENDIX II

A PRELAUNCH WIND CHECK PROCEDURE

INTRODUCTION

In the case of most R and D launches, and particularly with a manned payload, it is preferable to check the winds before launch and to delay the launch if wind effects appear excessive. This report is concerned with performing optimum wind checks, where optimum is defined as delaying a minimum number of launches while allowing no failures due to wind.

The winds of concern here are winds aloft - as contrasted with ground winds, which involve separate procedures and criteria for the launch decision. The prelaunch check of winds aloft involves measuring the wind

velocity profile and predicting the wind effect magnitudes for the flight. Both of these operations, and improved prediction techniques, are discussed.

PRELAUNCH WIND MEASUREMENT

The wind measurement technique presently used at the Atlantic and Pacific Missile Ranges uses balloons that rise at roughly 1000 feet per minute and will go to about 100,000 feet. The balloon carries a transmitter which allows it to be tracked and atmospheric data measured by sensors attached to the balloon to be telemetered. The most common equipment used presently for balloon position measurement is designated as GMD-1. Some improved equipment coming into service is designated as GMD-2. With the GMD-1 equipment, the balloon position is determined by using the elevation angle to the balloon (measured by the tracking antenna) and the balloon altitude (computed from the telemetered atmospheric data). With the GMD-2 equipment the position is determined from the elevation angle and the range (both measured by the tracking equipment). The balloon position is determined at regular intervals, and the wind velocity is found from the rate of change of position. The computed velocity is thus the average velocity over an altitude layer, (normally 1000 feet or more), and that velocity is assigned to the altitude at the center of the layer. The results of such a balloon sounding has been termed a wind profile in this report. Since it is evident that this is a smoothed profile, the profile information that is not detected has been referred to as profile detail. In regards to the time required to take a sounding and compute the profile, it presently requires about 1-1/2 to 2 hours to get velocities to an altitude of 60K feet.

Improved balloon sounding techniques are presently being developed which will afford more accuracy and will provide average wind velocities over smaller altitude layers. A promising method uses a high pressure, constant rise-rate balloon and the FPS-16 tracking radar. Also, computer facilities can be improved for more rapid data reduction and perhaps some optimum smoothing.

Table 1 shows a wind check schedule that is presently being used by the Aerospace Corporation. Note that wind predictions are used up to 12 hours before launch for making wind effect predictions, but actual sounding data is used after that time.

PREDICTION OF WIND EFFECTS

When any predicted wind effect magnitude exceeds its launch allowable the recommendation must be made to delay the launch. The prediction of the wind effect magnitudes is made on the basis of simulations through predicted and measured wind profiles and by means of the procedure that is outlined below.

Flight simulations are made using the T-2 day, T-1 day and T-12 hour predicted wind profiles, as soon as each profile is available. The wind effect magnitudes computed by the simulation are compared with the launch allowable values. In the case of an unsymmetrical vehicle the structural loading can be compared with the allowables by means of curves such as shown in

Table 1. A Wind Check Schedule

Meteorological Operation	Approx. Time Balloon is at Critical Altitude	Maximum Necessary Altitude	Wind Data Available	Method of Wind Effect Prediction	Wind Effect Predictions Completed
T-2 Day Prediction	-	60 K Ft	T-40 hrs	Simulation	T-36 hrs
T-1 Day Prediction	-	60 K Ft	T-16 hrs	Simulation	T-10 hrs
T-12 Hour Prediction	-	60 K Ft	T-7 hrs	Simulation	T-5 hrs
T-5-1/2* Hour Sounding	T-5 hours	60 K Ft	T-3-1/2 hrs	Simulation	T-2-1/2 hrs
T-2-1/2* Hour Sounding	T-2 hours	40 K Ft	T-1 hrs	Precomputed graph plus T-5 prediction	T-3/4 hrs
T-0* Hour Sounding	T + 1/2 hour	60 K Ft	T+6 hrs	Simulation	T+1 Day

* Times of balloon release

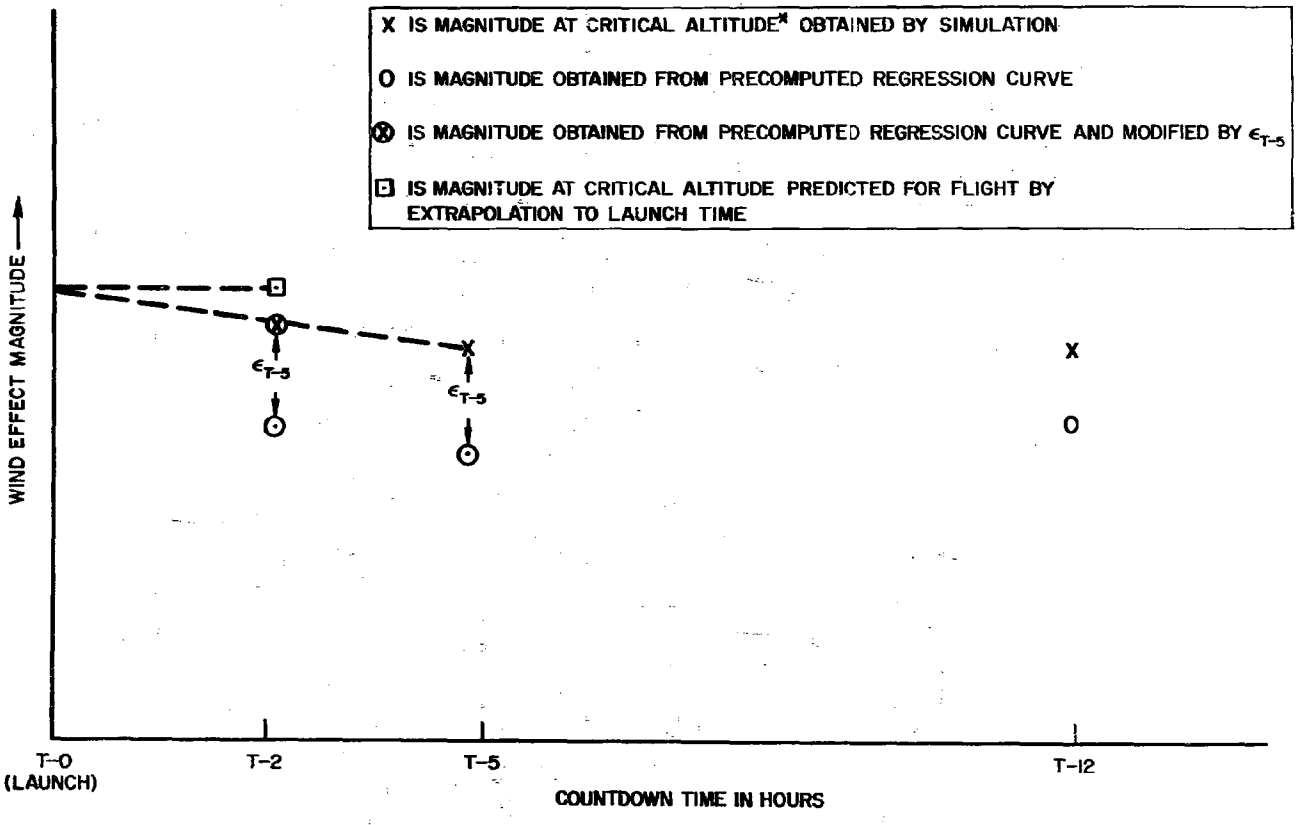
Figure 9. If a simulation shows that a predicted wind will cause the wind effect allowables to be exceeded, then an early decision to delay the flight could possibly be justified. However, such a decision would only be made after consultation with meteorological personnel, since the decision will certainly be influenced by the confidence assigned to such relatively long range predictions. In contrast, the decision regarding a delay in launch based on the complete procedure given here may be made less than an hour before launch time, so that the decision must be almost automatic. A short decision time is only workable if all parties concerned have reached prior agreement as to the criteria for decision.

During the last 6 hours or so before launch, a measured profile is usually the best prediction of the profile that will be seen in flight. Therefore, the T-5-1/2 hour and T-2-1/2 hour sounding data are used in making the final wind effect predictions.

When the data from the T-5-1/2 hour sounding is received, a simulation is made and the resulting wind effect magnitudes at the critical times-of-flight are plotted as "X's" at T-5 hours (the approximate time the balloon reached the altitude region of highest winds) on charts such as Figure 10. The critical time-of-flight for any wind effect considered here is the time at which its magnitude comes the closest to the allowable. In the case of structural loads, this critical time is found by plotting the computed loads versus time on a graph such as that of Figure 5 or Figure 9. (A graph is needed for each effect of interest.) Another set of wind effect magnitudes ("expected" maximum magnitudes) are obtained by taking the maximum wind from the sounding and then referring to regression curves of the type in Figure 8. These curves are usually obtained by simulating the missile through a number of random wind profiles, and they are meaningful because of the high correlation between wind velocity and wind effect magnitudes. These expected values are plotted as "0's", also at T-5 hours, and the differences ϵ_{T-5}, between the magnitudes from the simulation and the corresponding expected magnitudes are now computed.

No time is available to perform a simulation to compute wind effects using the T-2-1/2 hour sounding. However, the expected values are available from the regression curves. These values are then modified by ϵ_{T-5} to obtain more accurate values, and they are plotted as "X" at T-2 hours as illustrated in Figure 10. (The justification for the use of ϵ_{T-5} as a modifier is based on the assumption that the unique form of the profile that caused ϵ_{T-5} to have a value other than zero will tend to persist until flight time.) Final predicted values of wind effects can now be found by extrapolating to launch time the rate of change of magnitudes seen between the T-5-1/2 and T-2-1/2 hour sounding results. A final predicted value is shown in Figure 10 as "□", along with a simple geometric construction (dotted lines) to obtain it. This value can then be compared with the launch allowable value at the critical time-of-flight. If the predicted value exceeds the launch allowables, then the flight must be delayed.

Countdown holds which delay the launch often result from causes other than winds aloft. Such holds may necessitate a repetition of the previous wind sounding if (1) a total hold exceeding one hour occurs between the T-5-1/2



NOTE: T-5 AND T-2 ARE THE APPROXIMATE TIMES THAT THE SCHEDULED T-5 1/2 AND T-2 1/2 HOUR BALLOONS REACH THE ALTITUDES OF HIGHEST WINDS.

* CRITICAL ALTITUDE HERE REFERS TO THE ALTITUDE AT WHICH THE COMPUTED WIND EFFECT MAGNITUDE IS CLOSEST TO THE ALLOWABLE MAGNITUDE.

Figure 10. Graph for Launch Decision

and T-2-1/2 hour soundings, or (2) a total hold exceeding T-45 minutes occurs subsequent to the T-2-1/2 hour sounding. Therefore, there must be a capability to perform at least two extra soundings.

IMPROVED WIND EFFECT PREDICTION TECHNIQUES

An optimum wind check procedure has been defined as one causing a minimum number of launch delays while still preventing any in-flight failures. Thus to optimize, the launch allowable values must be increased to a point where the probability of failure when a launch is permitted starts to become appreciable. Although admittedly "appreciable probability" is difficult to define, there is certainly a practical limit to the conservatism that can be used in computing the launch allowables. Less conservatism is possible, of course, when more accurate information is available regarding the uncertainties. Better information regarding profile detail and wind variability (particularly for conditions of high wind velocity) would be especially helpful. Improved wind soundings, in the sense of measuring finer profile detail, will also be helpful, for then a smaller statistical allowance need be made for profile detail.

Predicting wind effects by actually simulating the missile through pre-launch soundings has the great disadvantage of requiring the services of trained men and computing facilities for a long period before launch. A much better procedure would be to have the launch criteria merely a function of wind parameters (velocity, direction, shear), so that, once the wind sounding data was available, a simple table or chart could provide information for the launch decision. This would be practical (1) if the missile was so insensitive to winds that extreme conservatism could be used in forming criteria without creating a seriously high probability of launch delay, or (2) if enough studies of the vehicle of interest could be conducted to compute accurate correlations between wind effects and a number of wind parameters.

PROBLEMS IN THE OXIDATION PROTECTION OF REFRACTORY METALS IN AEROSPACE APPLICATIONS

R. A. Perkins, L. A. Riedinger, and S. Sokolsky
Lockheed Missiles and Space Company
Sunnyvale, California

ABSTRACT

This paper presents an evaluation of protective coatings for refractory metals and critical problem areas in the light of requirements for radiation equilibrium cooled structures. Environmental conditions of temperature, pressure, flow, and time for hypersonic reentry vehicles are described. The effect of atmospheric pressure and flow on the performance of the most promising coating systems is discussed. Data are presented to show that rapid deterioration of silicide-base coatings can occur under reentry environmental conditions. Current oxidation test data can be misleading and potential uses of coated structures should be re-assessed. Improved coatings must be developed for use in low pressure environments.

Work Carried Out as Part of the Lockheed Independent Research Program

PROBLEMS IN THE OXIDATION PROTECTION OF REFRACTORY METALS IN AEROSPACE APPLICATIONS

R. A. Perkins, L. A. Riedinger, and S. Sokolsky
Lockheed Missiles and Space Company
Sunnyvale, California

INTRODUCTION

High-speed reentry into the earth's atmosphere results from various ballistic missile and aerospace missions. Radiation equilibrium cooling with heat shields and insulated structures provides an efficient and reliable method of thermal protection for structures of hypersonic reentry vehicles. Materials for radiation-cooled structures are selected for their ability to support the load and to resist the environment in which the structure functions. This must be done within particular limits of weight, fabricability, cost, and other considerations. The refractory metals, columbium, molybdenum, tantalum, tungsten, and their alloys, have the most promise for meeting the basic materials requirements.

Heat shields and structures capable of operating at temperatures to 4500° F can be designed and manufactured from existing materials. Refractory metal and alloy sheet are available in sufficient size and quality for most anticipated applications. Fabrication can be accomplished without undue difficulty and the basic shop technology for handling refractory metal sheet is developing rapidly. The problem of oxidation resistance and surface stability, however, presents a major deterrent to the effective use of these materials in aerospace vehicles.

Coatings are required to provide protection from oxidation and to increase emissivity for purposes of temperature control. During the minimum life period, oxidation must have no adverse effect on base metal properties. Targets for specific applications may vary considerably but, in general, radiation-equilibrium-cooled structures will require a life of 10 to 100 min at peak temperatures for one reentry.

Coatings which protect refractory metals from oxidation at temperatures to 3200° F are commercially available. The most promising coatings for molybdenum, columbium, tantalum, and tungsten are based on compounds of silicon (silicides) or aluminum (aluminides). Potentially, silicides represent the most important single class of materials for the oxidation protection of refractory metals. A number of advanced aerospace systems are being designed with heat shields and structural elements of silicide-coated refractory metals. Tests in air at 1 atm have indicated more than adequate protection at temperatures to 3000° F for several hours. Coatings which protect in air at temperatures above 3200° F have not been developed.

Coating performance and reliability under exact intended service conditions generally have not been determined. Present coatings have been evaluated by tests in slowly moving air or by exposure to high-temperature torches. The tacit assumption is made that air at 1 atm under different flow conditions represents the most severe environment. Most aerospace applications, however, involve exposure to static or

flowing air at reduced pressures. Since a gas phase is involved, pressure and flow may have a significant effect on both the thermodynamics and the kinetics of possible oxidation reactions. The fact that reduced air pressure has not been studied to any extent as a factor in the oxidation of coated refractory metals suggests a potential problem area in the use of these materials. The purpose of this paper is to re-evaluate the current status of protective coatings and to delineate critical problem areas in the light of specific environmental conditions and performance criteria for radiation-equilibrium-cooled structures.

ENVIRONMENTAL CONDITIONS AND DESIGN

The performance of materials in reentry structures is affected by the entire operational environment. All too often, temperature, per se, is looked upon as the controlling factor in materials selection and design. However, where oxidation resistance may be as much of a problem as strength and structural stability, gas pressure and dynamic flow must be given equal consideration with temperature. The precise relation of temperature, gas pressure, and dynamic flow as a function of time should be established for each component to serve as a basis for materials selection and design.

Thermal Environment

In general, the reentry velocity, with three basic aerodynamic parameters, establishes the precise thermal environment that must be accommodated.

(1) Lift-to-Drag Ratio at the Specific Mach Number (L/D). This factor is established by the basic vehicle design (shape) and the angle of attack (attitude). The upper limit of L/D is fixed by shape; the lower limit is determined by the attitude. The practical range of values for winged reentry vehicles is 0.2 to 2.0.

(2) Ballistic Reentry Factor ($W/C_D A$). This parameter is established by the shape and design of the vehicle and by the angle of attack. Practical values for winged vehicles are 30 to 200 lb/ft^2.

(3) Reentry Angle (γ_E). This defines the boundaries of a reentry corridor with the associated heating rate, total heat input, and dynamic pressure environments. The upper value determines the skip-out limit (the condition under which the vehicle will graze rather than re-enter the atmosphere) while the lower value is set by the maximum tolerance for heating and deceleration forces. Practical values range from -5 deg for the lower limit to -9 deg for the upper limit on super-orbital reentry.

The numerical value of these basic factors with the reentry velocity define the flight path or trajectory. Four partial differential equations of motion can be solved simultaneously to determine the precise altitude-velocity contour (2). The equations are solved at small integration intervals with the aid of a digital computer. A typical flight corridor of altitude-velocity contour plots for a light wing-loaded glider returning from orbit is shown in Figure 1. The upper corridor boundary is flown by maintaining the angle-of-attack of the vehicle such that its lift coefficient is always a maximum. The vehicle must be held at a high angle-of-attack to achieve this condition, and,

although deceleration occurs at the highest possible altitudes and heating levels are lowest, a larger portion of the vehicle (the entire lower surface) is exposed to higher heating than it would receive at lower angles-of-attack on trajectories with more severe heating. The lower corridor boundaries are selected on the basis of peak allowable leading-edge temperature and an arbitrary g-limit on the pilot.

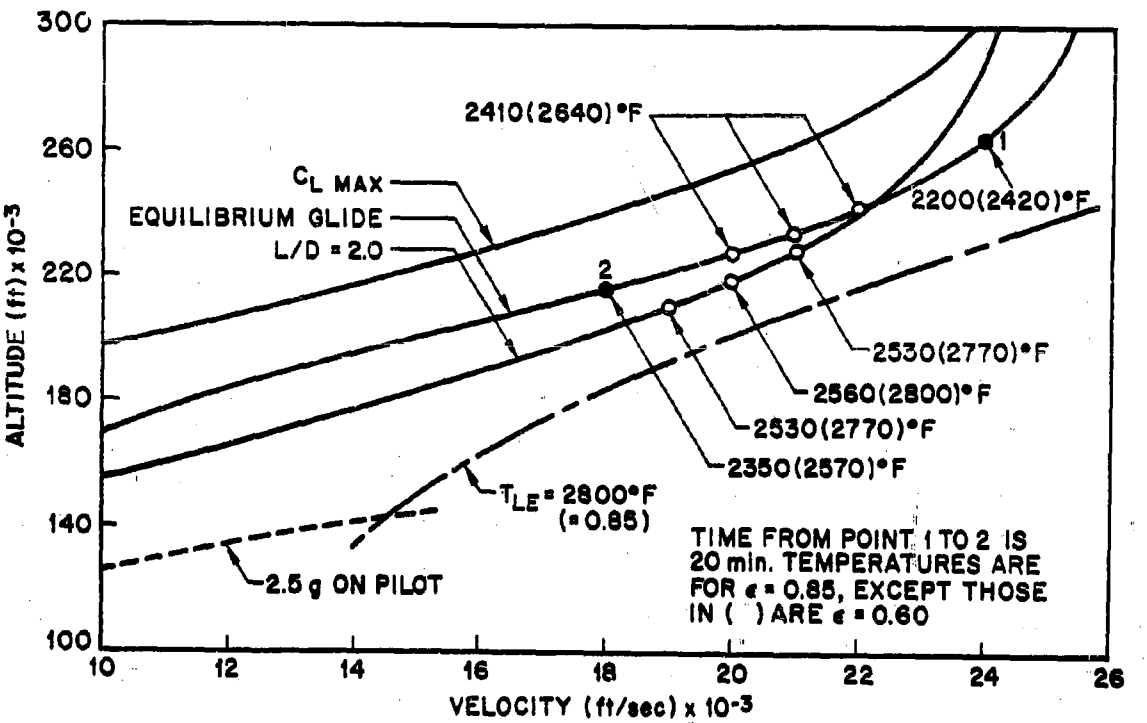


Fig. 1. Flight Corridor of Winged Reentry Vehicle (W/C_DA = 200 lb/ft^2)

Two intermediate trajectories are also shown in Figure 1. The so-called equilibrium glide trajectory results if one equates the sum of lift and centrifugal forces on the vehicle to vehicle weight, assuming all the forces to be colinear. Such a trajectory is not easily flown in practice, since it requires precise control of the vehicle to maintain the force balance as the vehicle decelerates. A constant lift-to-drag ratio trajectory is most easily maintained since it requires setting the vehicle at a constant and reasonable low angle-of-attack. Note that initially the trajectory in Figure 1 is steeper than the glide trajectory, but below 22,000 ft/sec, the two are quite similar. Temperatures along the constantL/D trajectory are higher than those on the glide trajectory, however.

From this plot, the rate of heat input can be calculated as a function of time. The heat input is expressed as an instantaneous heat flux ($\dot{q}$) in Btu/ft^2-sec, and a total heat input (Q) in Btu/ft^2. From a materials standpoint, the instantaneous heat flux and the duration of the heat pulse are the most important considerations. This determines the

optimum thermal protection system and the required materials characteristics. By combining the stagnation-point heat-transfer theory of Hoshizoki (3), the inviscid shock layer theory of Vinokur (4), a simple fit of equilibrium air data (5), and the best available experimental data for swept cylinders, the following expression for heat transfer (to a cold wall) at the stagnation line of a leading edge results:

$$\dot{q}_{L.E.} R_{L.E.}^{1/2} = \frac{873}{\sqrt{2}} (\bar{\rho}_{\infty})^{1/2} \bar{V}^{3.19} (\cos \Lambda') \tag{1}$$

At the enthalpy levels of significance in guide reentry, the assumption of a cold wall in computing the convective heat transfer results in slight conservatism in the results. Air density normalized to sea level (ρ_{∞}) and velocity normalized to 10,000 ft/sec ($\bar{V}$) are determined for any instant of time from the altitude-velocity contour plot. The effective sweep-back angle (Λ') can be computed from the sweep-back angle (Λ) by the following relation:

$$\Lambda' = \tan^{-1} \frac{\tan \Lambda \cos \alpha}{(1 + \tan^2 \Lambda \cos^2 \alpha)^{1/2}} \tag{2}$$

$$= \cos^{-1} (1 + \sin^2 \Lambda \cos^2 \alpha)^{1/2}$$

Equation (1) applies directly to the laminar convective heat transfer to a wing or other leading edge. For the special case of a nose cap, Equation (1) reduces to:

$$\dot{q}_{N.C.} R_{N.C.}^{1/2} = 873 (\bar{\rho}_{\infty})^{1/2} \bar{V}^{3.19} \tag{3}$$

The distribution of heat transfer from the stagnation line around the leading edge can be computed using the theory of Lees (6).

The so-called radiation equilibrium temperature can be calculated by means of the following simple relation:

$$\dot{q} = 0.478 \epsilon \left(\frac{T_W}{10^3}\right)^4 \tag{4}$$

This temperature (T_W) is defined as that reached by the surface if all heat transferred to the surface (in this case by convection) is re-radiated. It must be assumed that the skin of the leading edge is thin (zero thermal capacity) and that the emissivity of internal surfaces is small to satisfy this condition. Equation (4) can be used to calculate the surface temperature at any instant of time for any point on the structure for which the convective heat-transfer rate is known.

Reentry vehicles as such are divided into three major classifications based on velocity at the start of reentry: sub-orbital (18,000 to 24,000 ft/sec), orbital (24,000 to 26,000 ft/sec), and super-orbital (above 26,000 ft/sec) (Figure 2). The ballistic missile nose cone is the most familiar example of a sub-orbital vehicle. Peak heating rates ($\dot{q}$ = 1200 to 3000) are the highest for any class of vehicles, but the duration of the heat pulse is short. Ablative and heat sink materials are used for these structures and are not considered in the scope of this paper. Orbital vehicles, such as presently designed for manned and unmanned reentry capsules and boost-glide vehicles, sustain the lowest heat flux ($\dot{q}$ = 200 to 500) but over a much longer duration. The total heat input is similar to that for ballistic reentry depending on the range. Radiation cooling can be used effectively for the thermal protection of leading edges and adjacent structures. This represents one major area of use for coated refractory metals in aerospace vehicles. Super-orbital vehicles which re-enter the atmosphere on return from lunar or space missions present the most critical materials problem. Instantaneous heat fluxes on leading edges approach those for ballistic reentry ($\dot{q}$ = 200 to 1100) and they must be sustained for very long time periods. The total heat input (Q = 50,000 to 300,000) is highest for any class of vehicles. A combination of absorptive and radiative thermal protection systems is required.

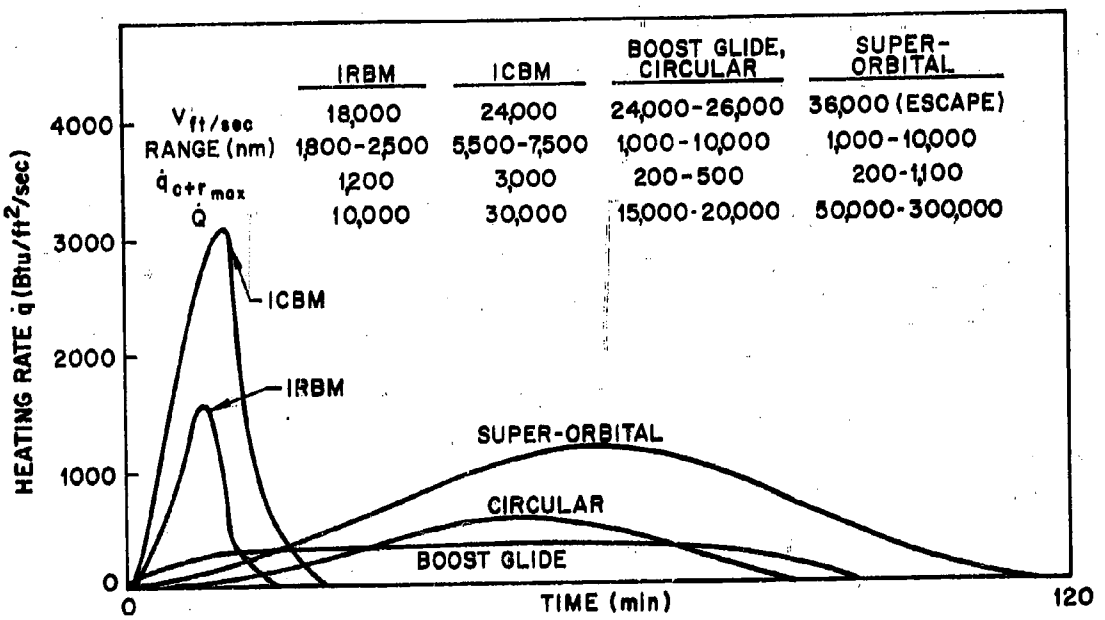


Fig. 2. Comparison of Typical Reentry Parameters

Various possibilities exist for the design and construction of leading edges and adjacent structures. The lightest design is provided by an uninsulted radiation-cooled structure. Such structures are limited to relatively low heat fluxes ($\dot{q} \simeq < 210$ Btu ft -sec) by the melting point of refractory metal components. However, higher heat fluxes can be sustained for short pulse heating cycles provided that thermal shock is not a problem. At a sustained heat flux of about 200 Btu/ft^2-sec, maximum temperatures approach 3000 to 5000° F, resulting in the need for refractory metals in heat shields and structures. The maximum heat rate as a function of reentry velocity can be controlled within certain limitations by varying the basic aerodynamic parameters. At a given L/D and γ_E increasing W/C_DA increases the heat rate and total heat input. At a given W/C_DA and γ_E increasing L/D decreases the heat rate but increases total heat input. Thus, the designer, to some extent, has the ability to match the thermal environment to the capability of materials systems. Increased materials capabilities are needed, however, to provide more latitude in design and flight path.

Pressure-Flow Environment

The critical peak heating period during atmospheric reentry occurs during descent from 400,000 to 100,000 ft. The corresponding static pressure of the atmosphere is 1.35×10^{-5} to 8.29 mm Hg. The temperature-pressure-time history at the stagnation point of a frontal structure for a light wing-loaded glider flying one possible reentry trajectory from outer space is shown in Figure 3. Pressure in this case is the static pressure of the atmosphere at any instant of time during descent. This pressure could exist on the inside of a vehicle assuming that openings to the atmosphere are remote from the stagnation region. Peak heating from 2700 to 4500° F occurs at a pressure of 8.5×10^{-3} mm Hg for a total time of 25.33 min. It has been shown that surface recession rates are negligible for tungsten under this condition and coating would not be required (7). This is not universally true for all reentry structures, but, in general, oxidation problems for internal surfaces and substructures are less severe than those for external surfaces.

The outer surfaces of a reentry vehicle are exposed to high-velocity air flow and to a much higher total pressure at any instant of time. The pressure at the stagnation line of the leading edge is easily computed, if it is assumed that the leading edge is unaffected by conditions at the vehicle's stagnation point (equivalent to assuming a swept leading edge of infinite extent). Upon making a number of simplifying assumptions regarding the flow field, the following simple, but accurate, expression for stagnation line pressure can be derived:

$$P_{SL} = \gamma_\infty \, M_\infty^2 \cos^2 \Lambda' \qquad (5)$$

The required pressure distribution is obtained by the artifice of matching a Newtonian distribution from the stagnation line past the sonic point to a Prandtl-Meyer expansion in the super-sonic region. The matchpoint is found as that point for which the respective pressure gradients are equal.

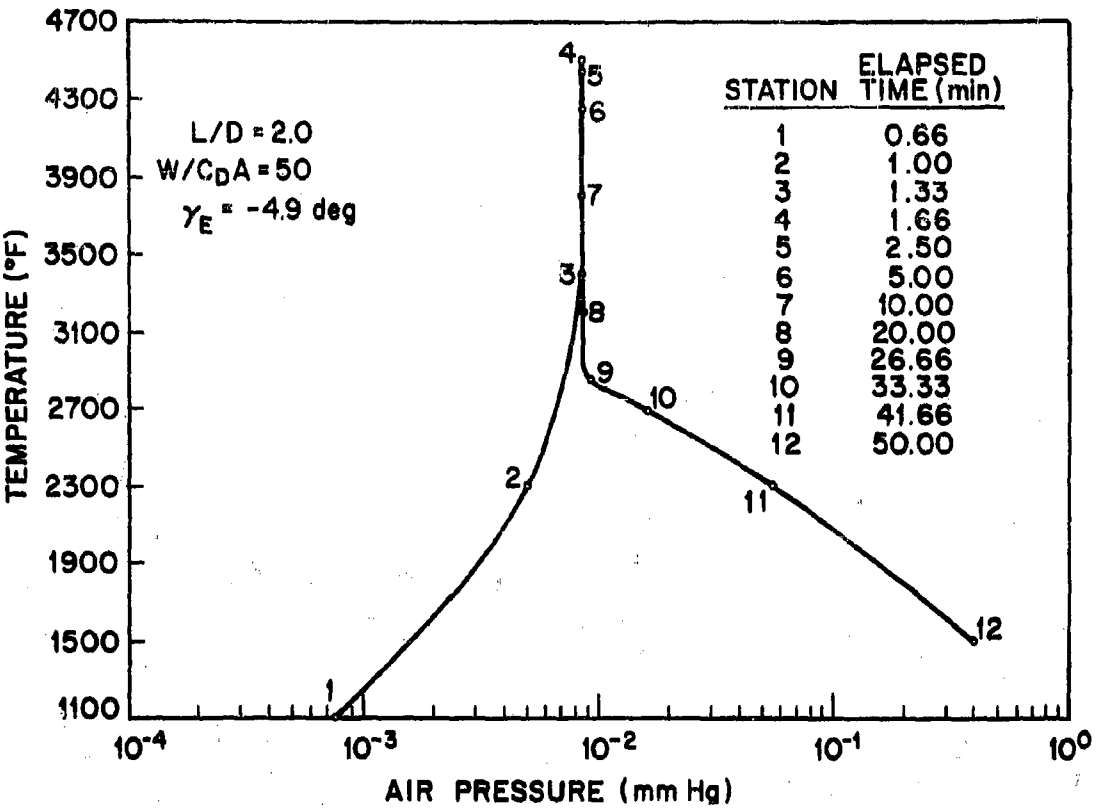


Fig. 3. Temperature – Pressure-Time History for Model Trajectory, Internal Metal Surface

The flight variation of temperature and pressure at the three locations on the wing leading edge on orbital reentry is shown in Figure 4. A total flight time of over 20 min is depicted by the curves. It is noteworthy that the temperature is nearly constant for approximately 11 min of this time, while the vehicle decelerates from 22,000 to 18,000 ft/sec. It is also significant that the pressure levels experienced by the leading edge are quite low. It may be surmised that if an unsealed joint is located just downstream of the leading edge, the pressure reached by the interior of the wing will be somewhere between that at the upper and lower surfaces, or less than 1 mm Hg during the time of peak heating.

SILICIDE COATINGS FOR REFRACTORY METALS

The most promising coatings for use at temperatures below 3200° F are based on compounds of the metal substrate with silicon. Of the 63 coatings for refractory metals under active development, 31 have a silicide base. Molybdenum as a substrate has received the most attention, and coatings based on molybdenum disilicide are in an advanced state of development. Typical properties for a variety of silicide coatings on molybdenum are given in Table 1. Tests in air at 1 atm indicate maximum life of

hundreds to thousands of hours at temperatures below 2700° F, ten to hundreds of hours from 3000 to 2700° F, and 0.5 to 10 hr above 3000° F. Similar coatings give adequate protection to columbium and tungsten.

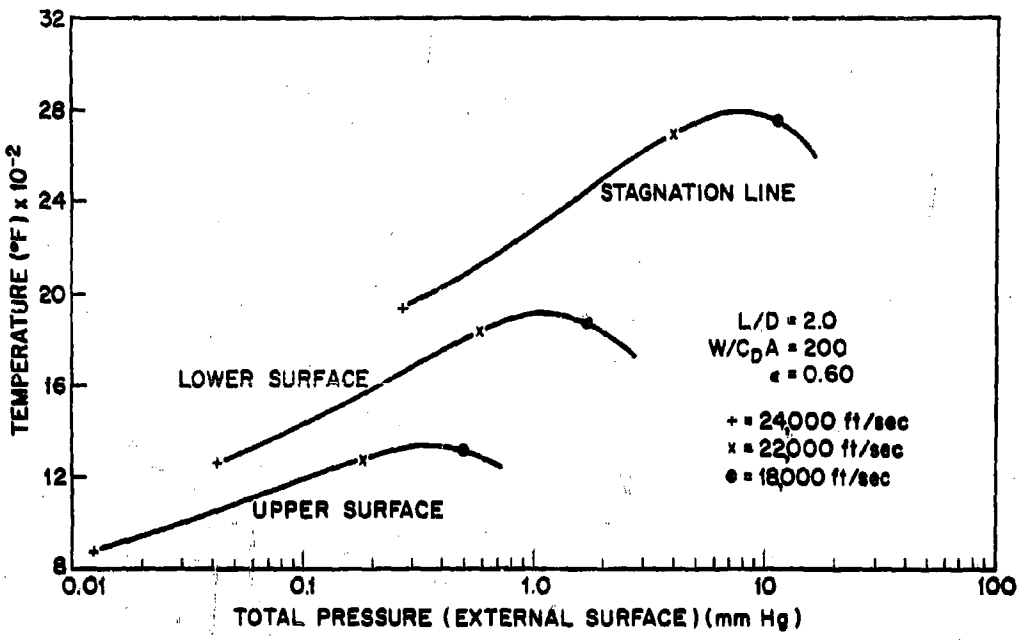


Fig. 4. Flight Variation of Temperature and Total Pressure

Table 1

MINIMUM LIFE OF $MoSi_2$ COATINGS FOR MOLYBDENUM [1]

TEMPERATURE		TARGET LIFE (Hr)	MEASURED LIFE (Hr)			
(°F)	(°C)		Disil (Boeing)	Durak-B (Chromizing)	PFR-6 (Pfaudler)	W-2 (Chromalloy)
2000	1093	100		>8500		
2400	1315	10		>225		20-30
2600	1427					10-20
2700	1482		15	>180		20-30
2800	1538				12.5	30-40
2900	1593				4.4	
3000	1649	3	<1	>20		
3100	1704				2.1	
3200	1760		<1		0.5	

Mechanism of Protection and Failure

Silicide coatings derive their protective nature from a thin coating of silica-base glass that forms during oxidation (8). The protective nature is intimately connected with the formation, growth, and stability of this glassy surface layer. Thermodynamic considerations suggest that a silica coating may not form or may be unstable in air at low pressure. Wagner (9) postulates that an active/passive transformation in the oxidation of silicon will occur as a function of oxygen pressure. Silicon forms two oxides, a stable solid (SiO_2) and an unstable gaseous oxide (SiO). At low oxygen partial pressures, solid SiO_2 will not be stable in contact with silicon at high temperatures. Gaseous silicon monoxide will form by the following reaction:

$$SiO_2 + Si = 2\,SiO(g) \qquad (6)$$

If the pressure of SiO gas at the surface (P_{SiO}^*) is less than the equilibrium pressure ($P_{SiO\,eq}$) for the above reaction at a given temperature, the surface will remain bare and a protective film of SiO_2 will not form. The rate of oxidation will be directly proportional to the amount of oxygen available for reaction. This is shown schematically in Figure 5. The transition from an active to a passive state will occur where P_{SiO}^* equals $P_{SiO\,eq}$. At this point, a solid oxide will form and the rate of oxidation will drop by several factors of ten. Pressure of SiO at the surface is shown to be a function of oxygen pressure and the diffusivity of gaseous species through a boundary layer. The pressure of oxygen required for the formation of solid SiO_2 on silicon at 2570° F was calculated to be 4.6 mm Hg. Experimental verification for the theory resulted in a measured pressure of 8 mm Hg. It was postulated further that the behavior would not be reversible and that silicon dioxide (SiO_2) films once formed would be stable to very low oxygen pressures.

This line of reasoning is directly applicable to oxidation behavior of compounds of silicon, i.e. $MoSi_2$ which are the major constituents of protective coatings. The equilibrium and reaction kinetics in this case would depend on the activity of silicon in the compound and that of silica in the oxide film. Since activity coefficients would be less than unity, the active-passive transition would be shifted to lower oxygen pressures or to higher temperatures. These considerations, however, strongly suggest that loss of silicon by selective oxidation at low pressure could occur within the intended range of temperatures and pressures for aerospace applications. Rapid deterioration of coatings and exposure of the substrate to oxidizing gases could result. Since diffusion of oxygen and silicon monoxide through boundary layers would be a rate-controlling factor, the high velocity on reentry could further accelerate coating destruction.

Of course, thermodynamic data give no indication of reaction kinetics which are the controlling factor in coating life. Extensive testing in static and moving air at low pressure is required to establish the role of pressure and flow.

Performance at Reduced Pressure and High Flow

The Martin Co. investigated the role of pressure on the performance of silicide coatings on molybdenum (10). No effect was found at pressures as low as 7.6 mm Hg. However, the results were inconclusive: time of coating failure could not be measured

accurately and exposure times at low pressure were limited to 1 to 10 min. Tests of commercial W-2 and Durak $MoSi_2$ coatings on molybdenum in hot flow and torch tests were made by NASA, Langley (11, 12). Investigators concluded that the stability of these coatings might be reduced under low-pressure environmental conditions encountered during reentry. Recently, Lockheed Missiles and Space Co. investigated the effect of pressure and flow on the life of commercially available pure and modified disilicide coatings on molybdenum.

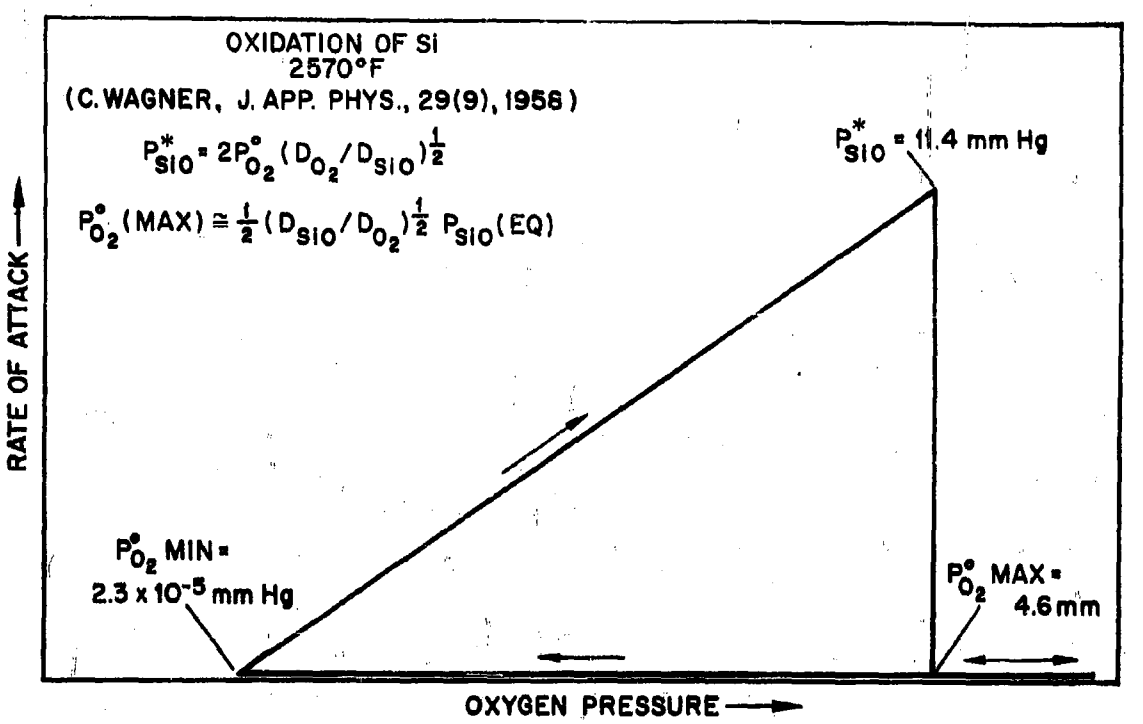


Fig. 5. Effect of Pressure on the Oxidation of Silicon

Samples of 0.10 to 0.125-in. diameter rods of unalloyed molybdenum were coated with advanced coating materials:

Vendor A – Pure $MoSi_2$, 0.0015 to 0.002-in. thick

Vendor B – Modified $MoSi_2$, proprietary composition, 0.0025-0.003-in. thick

These materials were selected to determine the effect of composition and structure on oxidation behavior in low-pressure environments. The samples were heated by direct resistance in a vacuum chamber with provisions for atmosphere and flow control. Temperature was measured optically with a micro optical pyrometer. Emissivity corrections for each coating were made by direct calibration against a heated bare molybdenum rod. Emissivity values for the coatings were similar and ranged from 0.6 to 0.7 in the as-received condition to 0.8 to 1.0 in the oxidized condition at temperatures from 2400 to 3200° F.

For each test, the maximum temperature for a 30 min. test without failure was determined. Where coatings failed in less than 30 min, tests were terminated to avoid complete destruction of the samples. Three methods for detection of failure were used: visual observation during test, measurement of the rate of oxygen consumption, and metallographic examination. The criteria for failure was localized (pinhole) or general attack of the coating with exposure of the substrate to the atmosphere. The results are summarized below:

		Max. Temperature for 30 min. Life ° F						
Coating	Thickness-mils	760 mm	160 mm	21 mm	5 mm	1.05 mm	0.21 mm	0.10 mm
Modified $MoSi_2$ (a)	2.8	3100	3075	3045	–	2925	2780	–
Pure $MoSi_2$ (a)	1.5	3200	–	3165	3150	2925	2780	2730
Pure $MoSi_2$ (b)	1.5	–	–	–	2750	–	–	–

(a) - static oxygen
(b) - flowing air

The coatings, as received, were found to contain numerous fine cracks and fissures on the surface. The structure consisted of columnar grains of $MoSi_2$ extending from the surface to a thin diffusion layer of unknown composition at the substrate. The modified coating contained, in addition, large amounts of a second phase concentrated in a band at about the centerline of the coating. On oxidation in air at 1 atm, a thick glassy oxide coating formed on the surface. The oxide healed the cracks and provided good protection from further oxidation. The maximum temperature in air as shown above was limited by the melting point of the coating (3250° F for pure and 3180 for modified $MoSi_2$). Failure in air occurred by blistering of the oxide film followed by rapid melting.

The initial tests at reduced pressure were conducted in static atmospheres of air or pure oxygen. The results of tests on the pure and modified silicides are shown in Figure 6. A slight decrease in the maximum temperature capability was noted with decreasing oxygen pressure to 5 mm Hg under static conditions. Failure occurred by blistering of the surface followed by rapid attack in a matter of minutes. The modified coating in air at 1 atm survived 30 min at 3100° F, while the sample failed in less than 1 min at a pressure of 160 mm Hg at the same temperature.

A very marked effect of pressure on coating performance was found at oxygen pressures below about 5 mm Hg. Although the glassy oxide appeared to form, it was very thin and no longer self-healing. Cracks and fissures broadened by oxidation and rapidly penetrated to the substrate. The maximum temperature for 30 min life without failure is reduced from 3100° F in air at 760 mm Hg to 2925° F in air or oxygen at 1.05 mm Hg and to 2780° F at 0.21 mm Hg. Metallographic examination indicated that samples were within a few minutes of failing at even lower temperatures. The curve for 45 to 60 min life without failure would be displaced to even lower temperatures at these pressure levels. Conversely, coating life at any given temperature is significantly reduced as pressure is decreased below 5 mm Hg. The coating which protected in oxygen at 21 mm Hg for 30 min at 3000° F failed in 7 min at 1.05 mm Hg and 3000° F.

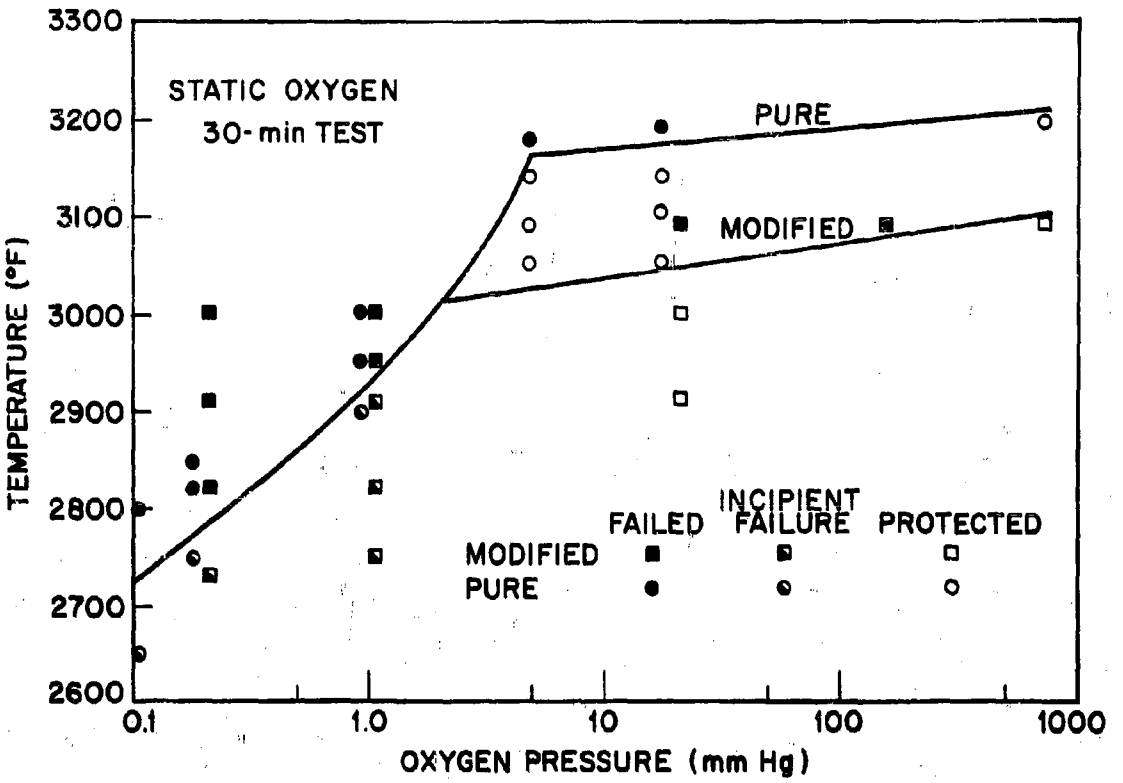


Fig. 6. Maximum Temperature For 30-Min Life of $MoSi_2$ in Static Air

The effect of pressure on coating performance was similar for both materials tested. The pure $MoSi_2$ coating withstood temperatures 100 to 120° F higher than the modified coating in static atmospheres at pressure from 5 to 760 mm Hg. This coating has a higher melting point and is more resistant to failure by blistering at temperatures of 3000 to 3200° F. At pressures below 5 mm Hg, both coatings have the same maximum temperature capability for 30 min life without failure. The maximum temperature for the pure $MoSi_2$ coating was reduced from 3200° F in air at 1 atm to 2730° F in oxygen at 0.10 mm Hg pressure.

The oxidation behavior at low pressure is illustrated further by studies of the rate at which oxygen reacts with the coating (Fig. 7). The initial rate of oxygen consumption is very high but drops within 2 to 3 min as the oxide layer forms. When a protective layer of silica has formed, the rate of oxygen consumption approaches zero. In these tests in static oxygen at 1.05 mm Hg pressure, the coating is protective at 2575° F. No reaction with oxygen was detected after 4 min of testing. At higher temperatures, the rate of consumption drops rapidly but still is significant after 30 min exposure. This indicates continued deterioration of the coating and is classed as an incipient failure. Although an oxide film appears to form, it no longer is protective and the coating suffers continued attack. With increased time of testing, to an hour or

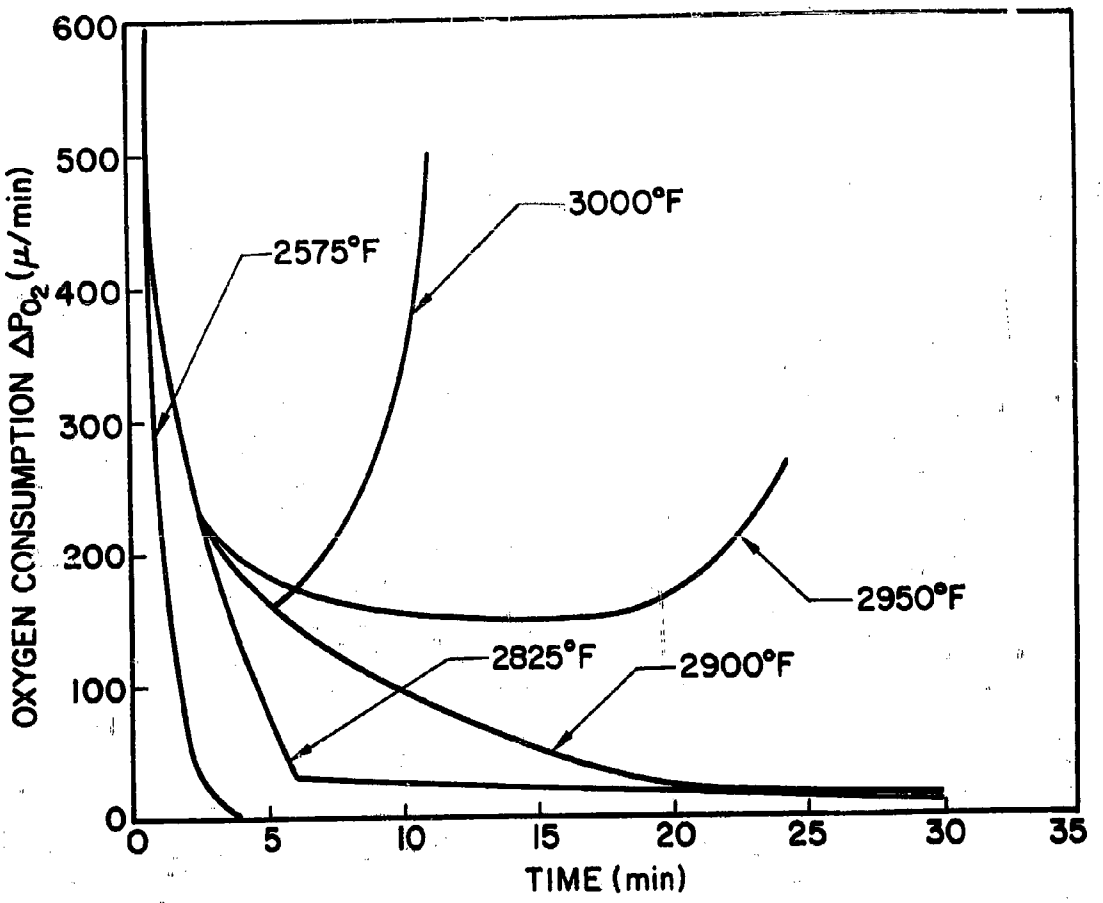


Fig. 7. Oxygen Consumption of $MoSi_2$ Coated Molybdenum in Oxygen and Air at 1.05 mm Hg

more, failure would occur. At still higher temperatures, an increase in oxygen consumption rate after the initial drop indicates failure of the coating and start of oxidation of the substrate. This technique has proved to be a particularly effective means of determining the time of failure in tests at low pressure. These data must be coupled with metallographic examinations however, to obtain a complete picture of coating behavior.

Visual and metallographic examination indicates that the mechanism and mode of failure at reduced pressure is the same for both coatings. Cracks and fissures are broadened by oxidation and rapidly penetrate to the substrate resulting in numerous pinhole (pit-type) failures (Figs. 8 and 9). Many internal voids develop and the coating fails rapidly from within. The pure $MoSi_2$ coating appeared to suffer more general surface attack whereas the modified coating suffered a more localized attack and penetration along cracks and fissures (Fig. 8). In all tests, the attack was localized and

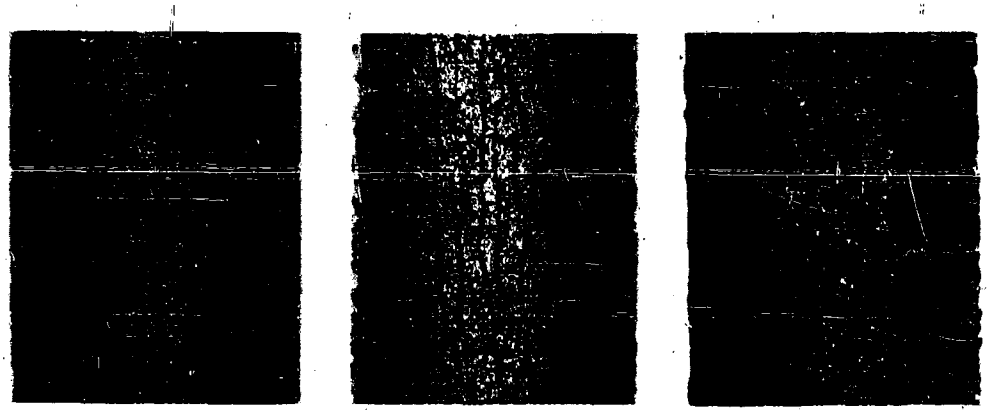


Fig. 8. Surface Appearance of Modified $MoSi_2$ Coated Molybdenum After Tests at 2900° F in Static Oxygen at Low Pressure (X30)

Fig. 9. Cross Section of Modified $MoSi_2$ Coated Molybdenum After Tests in Oxygen at Low Pressure (X150)

little change in total thickness of the coating as a whole was noted. At low pressure, failure proceeds along cracks, fissures, and grain boundaries.

All samples tested in oxygen above 5 mm Hg developed a smooth glassy oxide coating that appeared to be protective. At pressures below 5 mm Hg, the continuous glassy oxide coating did not form. Some solid oxidation products were noted, however, in many of the cracks and fissures. A gaseous oxidation product also was found in tests at low oxygen pressures. After tests, the cold walls of the reaction chamber were coated with a filmy white deposit. X-ray diffraction revealed this to be an amorphous material and spectrographic analysis revealed the major constituent to be silicon. Thus, it has been demonstrated that silicon is transferred from the sample to the furnace walls via the vapor phase. Vapor pressure and thermodynamic considerations suggest the most probable mechanism to be the formation of SiO as postulated by Wagner (9): $SiO_2 + Si = 2SiO(g)$. The silicon monoxide vapor is unstable and on cooling will dissociate into silicon and amorphous silica.

Since Wagner's theory predicted a hysterisis in oxide (SiO_2) stability, samples of $MoSi_2$ coated molybdenum were pre-oxidized in air at 2900 to 3000° F for 5 to 10 min to develop a glassy oxide coating. Subsequent tests in oxygen at low pressure (0.21 to 1.05 mm Hg) revealed no improvement in behavior. Oxygen consumption was drastically reduced in the initial 2 to 4 min but after that behavior was identical to that of samples that were not pre-oxidized.

Since the transport of both oxygen and silicon monoxide vapor appear to be involved in oxidation at low pressures, gaseous diffusion through boundary layers could be a rate controlling factor. Tests were made in high-velocity air jets at low pressure to determine the effect of gas dynamics on oxidation behavior. The specimen was placed transverse to the exit plane of a sonic orifice. The system was continuously evacuated with a mechanical pump while air was leaked through the orifice. The leak rate was balanced against the pump to achieve the desired static pressure. At a static pressure of 5 mm Hg, the gas velocity at the exit plane located 1-in. from the sample was calculated to be about 1000 ft/sec. Total (static plus dynamic) pressure at the surface of rod on the centerline of the jet was measured and found to be 10 to 13 mm Hg.

The results of tests of a pure $MoSi_2$ coating on molybdenum are shown in Figure 10. Under these conditions, the maximum temperature for a 30 min life was reduced from 3150° F in static oxygen at 5 mm Hg to 2750° F in moving air at 5 mm Hg. This coating which has a life in excess of 15 hr at 2700° F in air at 1 atm, failed in 130 min at 2700° F in moving air at 5 mm Hg pressure (Fig. 11). The maximum temperature capability in moving air parallels that determined by tests in static oxygen but is shifted to lower temperatures by 50 to 100° F at equivalent oxygen pressures. Thus, the rate of attack appears to be governed both by oxygen pressure and by the transport of gaseous species through boundary layers. Failures in moving air occurred over the entire heated surface and were not confined to the area on which the air jet impinged. This suggests that only a mild turbulence is needed to increase the transport of oxygen to the surface.

The equilibrium pressure of SiO from the reaction of silicon and silica as a function of temperature is shown in Figure 10. The data are consistent with the model proposed in that accelerated coating failure occurs when the total pressure, and hence the SiO pressure, is less than the equilibrium pressure of SiO for any temperature. Under these conditions, the reaction of silicon and silica to form silicon monoxide will be

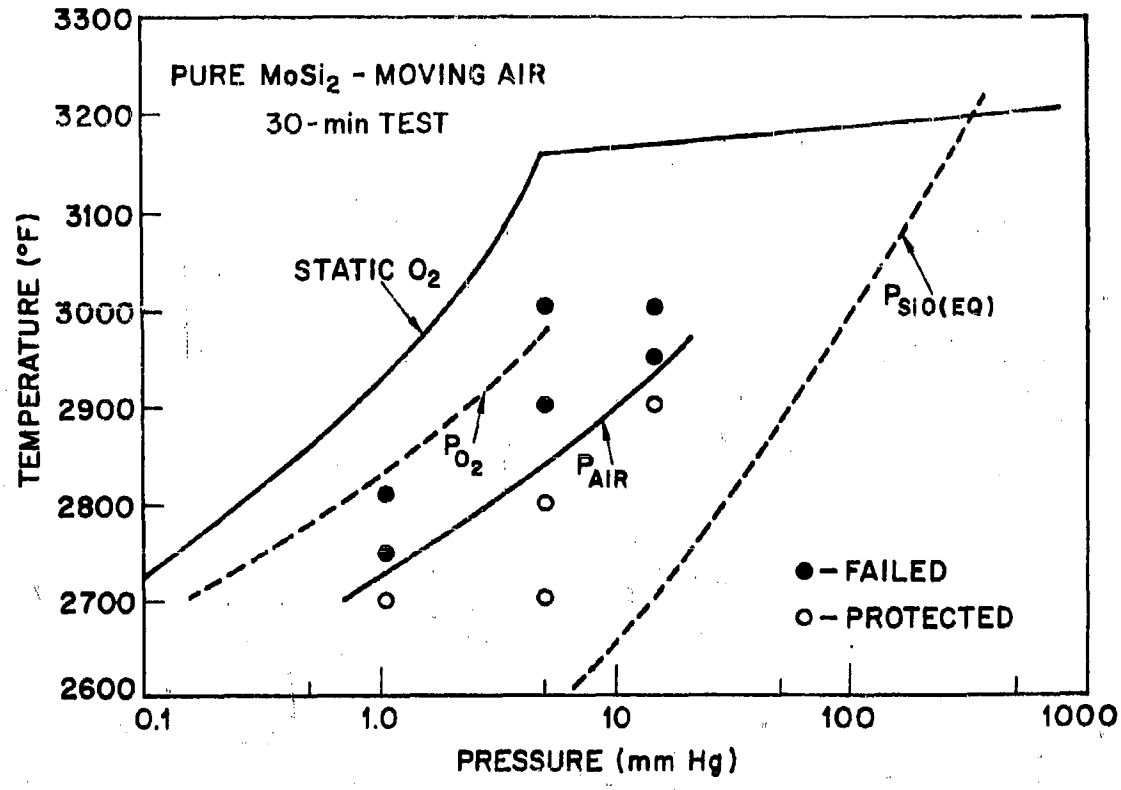


Fig. 10. Maximum Temperature for 30-Min Life of $MoSi_2$ in Moving Air

accelerated. Protective coatings of silica do not appear to form in static oxygen at pressures below 5 mm Hg and in moving air at pressures below about 50 mm Hg.

CONCLUSIONS AND RECOMMENDATIONS

1. Refractory metals for radiation-cooled equilibrium structures of reentry vehicles will operate at temperatures from 2000 to 4500° F in static and flowing air at pressures from 10^{-4} to 20 mm Hg. Time at temperature may range from 10 to 100 min for one reentry.

2. Based on tests in air at 1 atm, silicide base coatings will provide adequate protection from oxidation at temperatures to 3000° F. Coatings for protecting at higher temperatures must be developed.

3. The performance of silicide base coatings is significantly reduced in static oxygen at pressures below 5 mm Hg and in moving air at pressures below 50 mm Hg. The maximum temperature for a 30 min protective life may be reduced by as much as 500° F. Rapid deterioration of silicide base coatings can occur within the intended range of temperature-pressure-flow-time parameters for reentry structures.

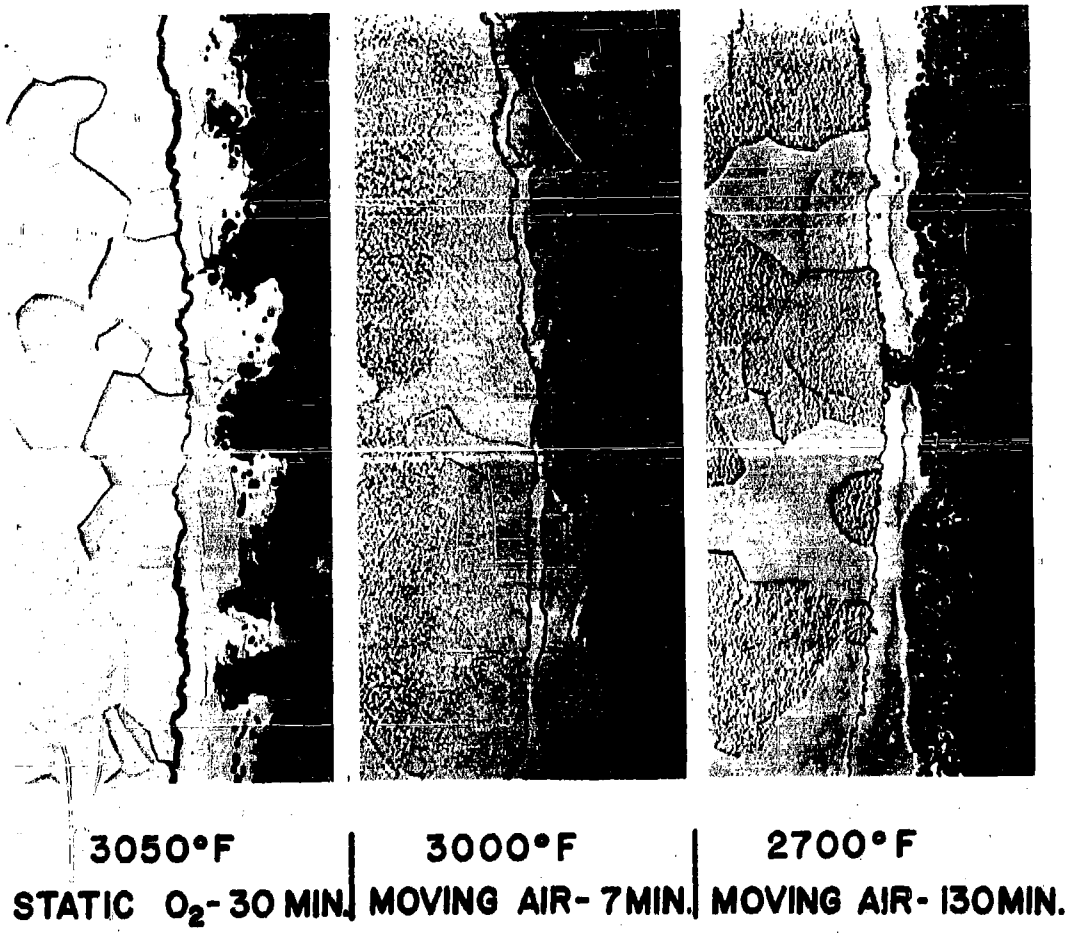


Fig. 11. Cross Section of Pure $MoSi_2$ Coated Molybdenum After Tests in Moving Air at 5 mm Hg (X150)

4. Failure at low pressure occurs by the formation of silicon monoxide, which is a gas at high temperature. Methods for improving oxidation resistance at low pressure without a reduction in maximum temperature capability must be found.

5. Current test and performance data on oxidation behavior are not directly applicable for predicting service behavior during atmospheric reentry or high-altitude flight. Data from tests in air at 1 atm can be misleading and should be used with caution. Potential uses of coated refractory metals should be thoroughly re-assessed.

6. Pressure and flow should be treated on an equal basis with temperature in the development and evaluation of coatings for aerospace applications. Coatings should be matched to the precise temperature pressure-time history of structures for a given vehicle and mission. These performance requirements can be calculated from basic aerodynamic parameters.

7. An extensive study of factors controlling performance of coated and uncoated materials in low-pressure environment should be undertaken.

SYMBOLS

A	Drag area (ft^2)
C_D	Drag coefficient
D	Drag force (LB)
M	Mach no.
p	Pressure (lb/ft^2)
$\dot{q}$	Heat transfer rate (Btu/Ft^2 sec.)
R	Radius of leading edge (ft)
T	Temperature (° R)
$\overline{V}$	Velocity normalized to 10000 ft/sec
W	Weight (lb)
γE	Reentry angle (deg)
α	Angle-of-attack (deg)
γ	Ratio of specific heats
ϵ	Emissivity
Λ	Sweep-back angle (deg)

Λ'	Effective sweep-back angle (deg)
$\bar{\rho}$	Density normalized to sea level

SUBSCRIPTS

N.C.	Nose cap stagnation line
L.E.	Leading edge stagnation line
∞	Free stream conditions
w	Wall conditions
S.L.	Stagnation line

REFERENCES

1. Report of the Subpanel on Coatings, Refractory Metals Sheet Rolling Panel, Materials Advisory Board, MAB-181, First Draft February 1961.

2. Hanshaw, C.E., Jacobsen, W.E., and G. Strasser, "Study of Structural Requirements of Re-Entry from Outer Space", WADD TR-60-886, May, 1961.

3. Hoshizaki, H., "Heat Transfer in Planetary Atmospheres at Super Satellite Speeds," A.R.S. Paper 2173-61, October 1961.

4. Vinokur, M., "Inviscid Hypersonic Flow Around Blunt Bodies," LMSC-48454, March 1959.

5. Bade, W.L., "Simple Analytical Approximation to the Equation of State of Dissociated Air," A.R.S. Journal, vol. 29, no. 4, April 1959.

6. Lees, L., "Laminer Heat Transfer Over Blunt Nosed Bodies at Hypersonic Flight Speeds," Jet Propulsion, vol. 26, no. 4, April 1956

7. Perkins, R.A., and Crooks, D.D., "Low-Pressure High-Temperature Oxidation of Tungsten," J. Metals, July 1961, p. 490.

8. Krier, C.A., "Coatings for the Protection of Refractory Metals from Oxidation," DMIC Report 162, November 1961.

9. Wagner, C., "Passivity During the Oxidation of Silicon at Elevated Temperatures", J. App. Phys., vol. 29, pp. 1295-7.

10. Wilks, C., and Magalotti, H., "Evaluation of Oxidation Resistant Coatings for Molybdenum," Contract NOw 60-0321c, The Martin Co., 15 June 1961.

11. Peters, R.W., and Rasnick, T.A., "Evaluation of Oxidation Resistant Coatings on Graphite and Molybdenum in Two-Arc Powered Facilities," NASA TN D-838, July 1961.

12. Trout, O. F., Jr., "Exploratory Investigation of Several Coated and Uncoated Metal, Refractory, and Graphite Models in a 3800° F Stagnation Temperature Air Jet," NASA TN D-190, February 1960.

MATHEMATICAL RELIABILITY TECHNIQUES

AN APPLICATION OF A NON-PARAMETRIC TECHNIQUE FOR RELIABILITY COMPARISONS

Kenneth M. Zenkere
Kenneth M. Hall

SYLVANIA ELECTRONIC SYSTEMS - WEST
A Division of Sylvania Electronic Products, Inc.
Reconnaissance Systems Laboratories
Mountain View, California

ABSTRACT

A frequent problem in the optimization of system effectiveness is to make objective decisions on operational procedures on the basis of limited a priori information. An analysis used to decide an effective maintenance policy for a complex subsystem is discussed in this paper.

AN APPLICATION OF A NON-PARAMETRIC TECHNIQUE FOR RELIABILITY COMPARISONS

Kenneth M. Zenkere
Kenneth M. Hall

SYLVANIA ELECTRONIC SYSTEMS - WEST
A Division of Sylvania Electronic Products, Inc.
Reconnaissance Systems Laboratories
Mountain View, California

1.0 INTRODUCTION

The basic configuration of the subsystem, involving partial sequential and parallel redundancy, is described and certain special problems arising from the increasing hazard rate (IHR) of one of the blocks are indicated. Two alternative maintenance procedures are considered; one in which a repair action is initiated when a redundant block fails, the other in which failures in the redundant blocks are repaired at scheduled times and preventative maintenance is performed on the IHR block.

Models for predicting the mean time between failures (MTBF) of the subsystem for each of these alternates are derived. From these models the relationship between the subsystem MTBF and the failure distribution of the IHR block is determined.

A non-parametric method for predicting certain limiting characteristics of the IHR failure distribution on the basis of a small sample censored test is discussed and then applied to predict which of the alternative maintenance procedures can be expected to give the greatest subsystem MTBF.

2.0 SUBSYSTEM DESCRIPTION

In this section a reliability description of the subsystem is given. Since the specific functions involved are being discussed from the reliability point of view, no operational discussion of the functions will be included. For reliability purposes, we may distinguish three types of functions within the basic subsystem.

2.1 SERIES FUNCTIONS

One basic group of equipment can be characterized by having an exponential distribution of time to failure and by the fact that its functions preclude any practical form of redundancy. The equipment involved in this set of series functions, which we will call set A, may be regarded as setting a basic limit on the subsystem MTBF.

2.2 PARALLEL FUNCTIONS

Another group of equipment can be characterized by exponential failure distributions and the fact that individual units can be utilized in a parallel redundant configuration. This equipment will be called set B.

2.3 SEQUENTIAL FUNCTIONS

The final function to be discussed is the most critical one from the reliability standpoint. The first characteristic of this function is that the failures are dominated by those of a single vacuum tube which, on the basis of available information, may be presumed to have a non-exponential distribution of time to failure, although the precise nature of the distribution is unknown.

The second significant observation concerns the fact that the function cannot be operated in a parallel redundant manner, although some forms of sequential redundancy are technically feasible. A third factor entering the picture is that any down time during a switching operation from one tube to another must be shorter than the normal tube warm-up time.

The equipment involved in this final function will be called set C.

2.4 SUBSYSTEM CONFIGURATION

A preliminary analysis of various alternatives indicated that, in order to fulfill the reliability and operational requirements, the equipment configuration should be that given below.

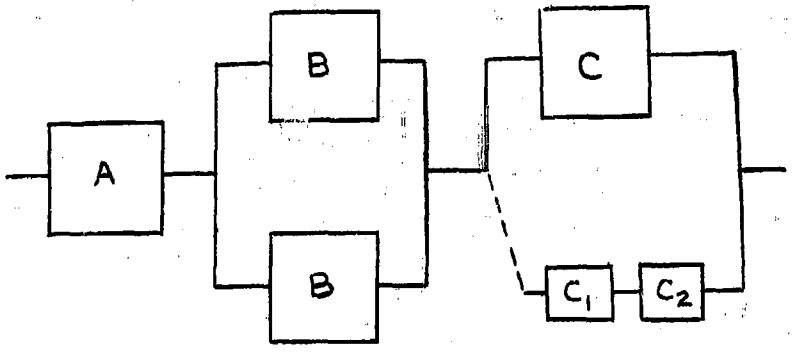


Fig. 1. Reliability Configuration

The subsystem will fail if block A (the series function) fails or if both of blocks B fail. The subsystem will fail if C_1 (logic and switching circuitry) fails and then C (the operating tube and associated circuitry) fails. The subsystem will also fail if block C fails and then block C_2 (then operating at full power) fails. The subsystem does not fail if C_1 fails after C has failed. Any failure within the equipment, whether it causes a subsystem

failure or not, is immediately indicated at a status board, so that the "failure state" of the equipment is always known.

3.0 MODELS FOR PREDICTION

Any satisfactory reliability model for the configuration described above must take into account the possibility that repairs may be completed before the failures have accumulated in such a way as to cause subsystem failure.

Another factor that should be evaluated arises from the presumed IHR nature of the vacuum tube failure distribution. By considering the hazard rate (which may be defined as the probability the tube will fail in a small time interval, given it has survived until that time) as an increasing function of time, we are simply stating that a "wearout" phenomena is involved. By replacing such a device at a time well before it may be expected to fail, the probability of failure in any interval may be sharply reduced.

Two alternates are discussed here. In both of these alternates, a subsystem failure will be repaired as rapidly as possible. In the first alternate, a failure in the redundant circuitry that does not cause a system failure will initiate a repair action to be completed within a few days. Parts will be replaced only when they fail. In the second alternate, failures that do not cause subsystem failure will be repaired at a prescheduled time and, at that time, the vacuum tubes will be replaced regardless of whether or not they or any portion of the equipment, have failed. The scheduled maintenance (or periodic replacement) interval would be on the order of 3 to 6 months.

A brief outline of the model derivations for these alternatives is presented below.

3.1 MODEL FOR REPAIR-ON-FAILURE POLICY

If the time to complete a routine repair is exponentially distributed, a repair rate, μ repairs per hour, may be established by solving

$$\mu = \frac{-\ln(1-P_R)}{T} \qquad (1)$$

where P_R is the probability that the repair will be completed within T hours. If $P_R = 0.95$ and $T = 168$ (i.e., a repair will be completed within one week with probability 0.95), μ becomes 0.01783.

While the exponential assumption may appear to be questionable, it can be justified by noting that the number of times a crew will become available for the repair in any time interval can be considered as a Poisson distribute and that the major portion of the repair time will be spent in waiting for a crew to become available.

If the state of the subsystem is described by specifying which blocks are "down", the probability of entering any such state in a single step is a

function only of the present state of the subsystem. Accordingly, the process may be described by a finite Markov chain.* There are four significant states of the subsystem for which there is no loss of subsystem function.

State 1: Nothing down.

State 2: Either of the partially sequential functions down.

State 3: Either of the "B" blocks down.

State 4: Either of the "B" blocks and either of the partially sequential blocks down.

It should be noted that another state of satisfactory performance, "Logic and Switching" failure after operating tube failure, is not included here, since the event is of very small probability in the short repair period.

Let us define a 4 x 4 matrix, Q, whose entries, q_{ij}, give the probabilities of going from state i to state j in a single step. It can be shown that

$$\left[I - Q\right]_1^{-1} \left[M\right] = \Theta_s$$

where I is the identity matrix, M is a column vector of the mean holding times in each state and Θ_s is the MTBF of the subsystem**.
The post subscript 1 means to use row 1 of the matrix $\left[I - Q\right]^{-1}$. It is important to note here that if the tubes are replaced only when they fail, the distribution of the time between tube failures in the system (i.e., in the entire complex) will ultimately approach an exponential distribution.

Defining (in appropriate units),

λ_A = failure rate of Block A

λ_B = failure rate of Block B

λ_C = failure rate of Block C

λ'_C = failure rate of Blocks C_1 and C_2

μ = repair rate of the subsystem

$B_1 = \lambda_A + \lambda_C + \lambda'_C$

---

* Finite Markov Chains, Kemeny and Snell; Van Nostrand, 1960, p.25.

**Use of Finite Markov Chains in the Reliability Analysis of Systems, Hall and Zenkere, Proceedings of the 1962 National Winter Convention on Military Electronics.

$$B_2 = \lambda_A + 2\lambda_B + \lambda_C$$
$$B_3 = \lambda_A + \lambda_B + \lambda_C + \lambda'_C$$
$$B_4 = \lambda_A + \lambda_B + \lambda_C$$

and carrying out the indicated operations we obtain

$$\theta_S = \frac{1}{2\lambda_B + B_1} \cdot \frac{1 + H}{1 - \frac{\mu H}{2\lambda_B + B_1}} \tag{2}$$

where

$$H = \frac{\lambda_C + \lambda'_C}{\mu + B_2} + \frac{2\lambda_B}{\mu + B_3} + \frac{2\lambda_B}{\mu + B_4} \cdot \left(\frac{B_1}{\mu + B_3} + \frac{\lambda_C + \lambda'_C}{\mu + B_2} \right)$$

After substituting the predicted failure rates and the value of μ computed from Equation 1 and then simplifying we obtain

$$\theta_S = \frac{5.816\,\theta_T^4 + 1456\theta_T^3 + 113600\theta_T^2 + 2812000\theta_T}{0.0003357\theta_T^4 + 0.09444\theta_T^3 + 413.4\theta_T^2 + 50860\theta_T + 1562000} \tag{3}$$

where θ_T is the MTBF of the tube. The MTBF of the tube on standby is assumed to be $4\theta_T$. A graph of θ_S as a function of θ_T (in the range of interest) is given in Figure 2.

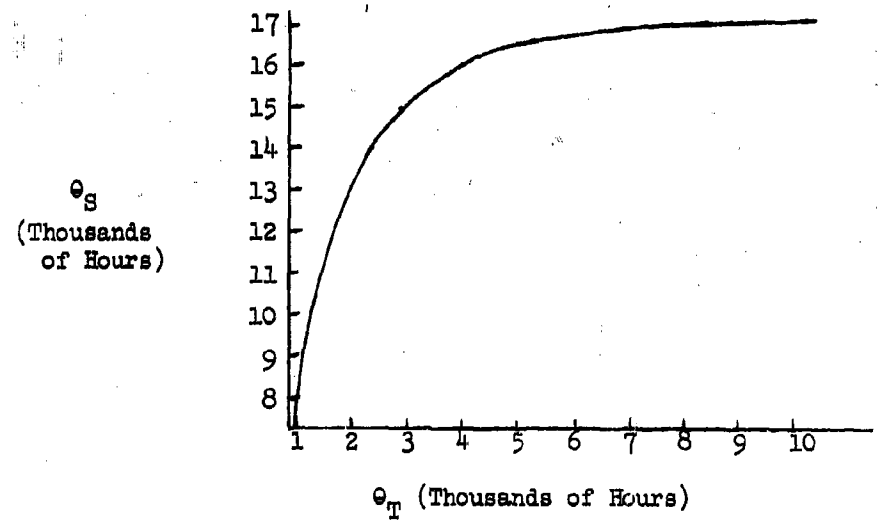


Fig. 2. Graph of θ_S as a Function of θ_T

3.2 MODEL FOR PERIODIC MAINTENANCE POLICY

To derive a model for predicting the subsystem MTBF under the alternate policy a somewhat different approach is indicated.

To begin, let us simplify the analysis by assuming that an effective failure rate of the tube can be specified by the equation

$$\lambda_e = \frac{-\ln R_T(k)}{k} \tag{4}$$

In this equation λ_e(failures per hour) is the effective failure rate of the tube, $R_T(t)$ is the reliability function of the tube, and k (in hours) is the interval between tube replacements. An exponential reliability function, having this effective failure rate, is a sharp lower bound of $R_T(t)$ in the interval $o \leq t \leq k$, as long as the hazard rate of the tube is non-decreasing everywhere within this interval*. If k is small compared to θ_T (the MTBF of the tube) the effective reliability function of the tube, $R_e(t) = \exp(-\lambda_e t)$ is very nearly equal to $R_T(t)$ in the interval.

With this single simplifying assumption we can easily derive an expression for the MTBF of the subsystem when the tubes are replaced at intervals of k hours and any necessary repairs to the redundant blocks are performed at the same time.

By first finding the reliability functions for each of the separate function sets (A, B, and C of section 2) for the configuration in Figure 1 and then taking the product of these reliability functions, we find the reliability function, $R_S(t)$, for the system without repair or tube replacement. This function is valid for $t \leq k$. The subsystem MTBF with periodic repair and tube replacement is then

$$\theta_S = \frac{\int_o^k R_S(t)dt}{1 - R_S(k)} \tag{5}$$

where

$$R_S(t) = R_A(t) \cdot R_B(t) \cdot R_C(t) \tag{5a}$$

$$R_A(t) = \exp(-\lambda_A t) \tag{5b}$$

*This bound was called to the authors' attention by Dr. R. Barlow of the G.T. and E. Laboratories in Menlo Park, California.

$$R_B(t) = 2\exp(-\lambda_B t) - \exp(-2\lambda_B t) \text{ and} \tag{5c}$$

$$R_C(t) = \frac{\lambda_{C_1}}{\lambda_{C_1} + \frac{\lambda_e}{4}} \cdot \exp\left[-\left(\lambda_{C_1} + \frac{5\lambda_e}{4} + \lambda_p\right)t\right]$$

$$-\left[\frac{\lambda_e + \lambda_p}{\frac{7\lambda_e}{4} + \lambda_p}\right] \cdot \exp\left[-\left(\lambda_{C_1} + \lambda_e + 2\lambda_p\right)t\right]$$

$$+\left[\frac{\lambda_e/4}{\lambda_{C_1} + \frac{\lambda_e}{4}} + \frac{\lambda_{C_1}}{\lambda_{C_1} - \frac{3}{4}\lambda_e - \lambda_p}\right] \cdot \exp\left[-\left(\lambda_e + \lambda_p\right)t\right]$$

$$+\left[\frac{\lambda_e + \lambda_p}{\frac{7}{4}\lambda_e + \lambda_p} - \frac{\lambda_{C_1}}{\lambda_{C_1} - \frac{3}{4}\lambda_e - \lambda_p}\right] \cdot \exp\left[-\left(\lambda_{C_1} + \frac{\lambda_e}{4}\right)t\right] \tag{5d}$$

where(in appropriate units),

λ_A = failure rate of Block A

λ_B = failure rate of Block B

λ_e = failure rate of the tube defined in Equation 4

λ_p = failure rate of the remaining circuitry in Block C

λ_{C_1} = failure rate of the logic and switching circuitry.

(It is assumed that the effective failure rate of the tube on standby is one fourth that of the operating tube.)

Equation 5d is complex in detail, but simple in principle. The only difficulties to be overcome are in remembering that "survival" is sometimes dependent on the order in which the failures can occur. By listing all independent conditions of survival and analyzing each in a straightforward manner, Equation 5d was derived.

When we substitute k = 2200 (the quarterly replacement interval) into Equation 5 an interesting fact develops. Even if the tube never fails (i.e., $\lambda_e = 0$) before being replaced, quarterly maintenance of the subsystem yields a lower value of the subsystem MTBF than the first policy whenever Θ_T exceeds

3700 hours. A tube that never failed before 2200 hours and had a MTBF of less than 3700 hours seems to be unlikely, but it is possible to specify an upper bound to the probability of the quarterly replacement improving the subsystem MTBF. An analysis is discussed in subsequent sections.

4.0 TUBE CHARACTERISTICS

Twenty-six tubes were tested for 1000 hours and exactly one of them failed. By regarding "survival" for 1000 hours as a binomial event with probability θ, it can be shown that the fiducial probability of a confidence interval $(\underline{\theta},\overline{\theta})$ containing θ is

$$P_2 - P_1 = (\overline{\theta})^{26} - \underline{\theta}^{26}\,(26-25\,\underline{\theta}) \qquad (6)^*$$

4.1 IHR ANALYSIS

For periodic replacement to be effective, the hazard rate must increase with time. Since the actual form of the life distribution is not known, the IHR (increasing hazard rate) assumption was made in order to make inferences on the characteristics of the actual distribution.

If the actual modal life of a tube is less than 2200 hours, then quarterly replacement does not buy anything because replacement occurs after the time at which failures are expected. (The proof of this statement follows directly from IHR considerations). On the other hand, from section 3.2 we deduced that if the expected life of a tube is greater than 3700 hours, then the replace-upon-failure policy is superior. Consequently, in order for quarterly replacement to show an advantage over the alternative policy, the modal life of a tube must necessarily lie between 2200 and 3700 hours.

Past experience has shown that for similar tubes, the truncated Normal (or Gaussion) distribution affords a satisfactory model for describing the life of the tubes. Under the assumption that tube life is distributed with a truncated Normal with parameters μ and σ (both μ and σ unknown), it is then necessary to determine whether the Normal distribution is favorable for the quarterly replacement policy.

By this, we mean that the actual distribution must have a mode which lies between 2200 and 3700 hours if quarterly replacement is to be advantageous. We desire to make a confidence statement about the possible existence of such a distribution which might give an advantage to a quarterly replacement policy.

If F denotes the Normal distribution function, then the survival probability, x, at time t is given by

$$x = F\left(\frac{t-\mu}{\sigma}\right) / F\left(\frac{\mu}{\sigma}\right)$$

*Hald; Statistical Theory with Engineering Applications; p. 697F.

Here, μ must be interpreted to be the mode since we are working with a Normal distribution truncated at the origin (t=o). This accounts for $F(\frac{\mu}{\sigma})$ in the denominator.

Since t = 1000 hours (the test length) we obtain

$$x = F\left(\frac{1000-\mu}{\sigma}\right) \Big/ F\left(\frac{\mu}{\sigma}\right)$$

$$= F\left[\frac{\mu}{\sigma}\left(\frac{1000-\mu}{\mu}\right)\right] \Big/ F\left(\frac{\mu}{\sigma}\right)$$

Substituting 2200 and 3700 for μ, we obtain $\underline{x}$ and $\bar{x}$, respectively, i.e., if we define $v = \sigma/\mu$

$$\underline{x} = F\left(\frac{6}{11V}\right) / F\left(\frac{1}{V}\right)$$

$$\bar{x} = F\left(\frac{27}{37V}\right) / F\left(\frac{1}{V}\right)$$

Note that $\underline{x} \leq \bar{x}$ for any V.

We can now utilize equation 6 to obtain

$$P_2 - P_1 = \bar{x}^{26} - \underline{x}^{25}(26-25\underline{x})$$

which is graphed in Figure 3 as a function of V. From Figure 3 we find the maximum value of $P_2 - P_1$ to be 0.11*. We can then conclude, it is not likely that the life distribution is favorable for the quarterly replacement policy**.

5.0 CONCLUSIONS

Essentially, it has been shown that we can draw some strong inferences on the basis of extremely limited test data provided we can justify certain weak assumptions (IHR in this case) about the nature of the parent population.

By analyzing the allowable family of truncated Normal distributions in more detail the fiducial probability that improvement is possible would

---

* Experience has shown that for tubes of this type, usually $0.3 \leq V \leq 1.0$.

** $P_2 - P_1$ is an upper bound to our confidence of improvement since a distribution with $2200 \leq \mu \leq 3700$ hours will not necessarily give an advantage to the quarterly replacement.

become even smaller. Doing this would require tedious computations which are not justifiable, since the simple analysis above allows a satisfactory inference.

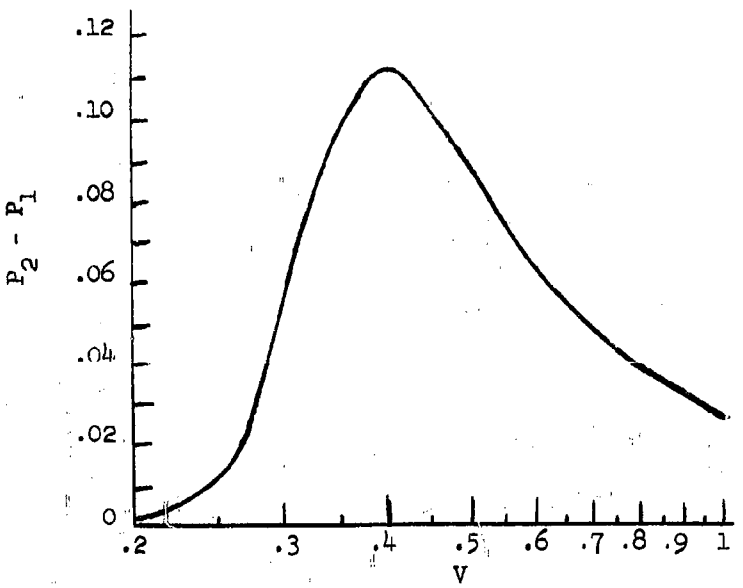


Fig. 3. Graph of P_2-P_1 as a Function of V

Failure Prediction for Nonmaintainable Electronic Components

C. W. G. Fulcher
General Electric Company
Defense Systems Department
King of Prussia, Pa.

ABSTRACT

The properties of the absorbing Markov chain were used to predict times to failure for nonmaintainable electronic components, the object being to identify defective components which fail before the end of design life although acceptance test performance is within design tolerances. Predicted failure times for several prototype electronic elements are compared to actual times.

Failure Prediction for Nonmaintainable Electronic Components

C. W. G. Fulcher
General Electric Company
Defense Systems Department
King of Prussia, Pa.

INTRODUCTION

A problem of interest to the missile and space industry is the timely identification of defective components which appear to operate satisfactorily during preliminary inspections but which fail prematurely in service. The problem becomes more difficult when life-test data are not available, the systems are nonmaintainable, recovery and inspection are not feasible after component failure, and design life is moderately long (10,000 - 30,000 hours). Consequently, the identification of components which will fail in service prior to the completion of intended design life would be desirable. Such identification would be limited to components whose levels of performance degrade to unacceptable levels before the end of design life and would not include identification of components which would, perhaps, fail after exposure to unusual environmental conditions.

From standpoints of economy and convenience, it would be desirable to use data accumulated during the customer's acceptance tests to identify those elements which operate within design tolerances during the acceptance tests, but which fail before the end of design life. If such test data were used, it would limit the detection of impending failure to those components whose reduction in performance could be revealed by the particular measurements being made. The importance of using measurements which are sensitive to performance degradation is thus established. Service failures caused by unacceptable reductions in performance of passive or active elements (whose performance variations cannot be detected by the recorded measurements) may be relegated to chance. Duplication of the expected service environment during acceptance tests is desirable for performance analysis and is assumed to exist. Such analysis of performance trends is not intended to replace but to supplement current quality control or reliability procedures in the identification

of defective units through greater use of available test data.

To identify nonmaintainable components which fail prematurely, a mathematical model was formulated utilizing the theory of Markov chains. The model was then checked using life-test data from long-life electronic elements. Design life was defined as being 20 time intervals long. The time required for components to fail, when failure occurred before the lapse of 20 intervals, was adequately predicted at the end of two test intervals. Components which did not fail in 20 intervals were identified as good components, with one exception, at the end of the second interval.

To obtain the mathematical model, the equipment for which the model is applicable is assumed to have a type of performance that generally degrades with passage of time. The model can be required to recognize failure at any given level, or levels, of performance. Once the equipment has failed, the assumption is made that it is removed from service. A fundamental probability matrix is constructed using data from tests conducted on a small number of prototype electronic components whose performance satisfied the above assumptions. Various quantities are obtained from matrix algebra with the aforesaid results.

RESULTS

Governing Equations

The expected failure time for a component may be expressed as

$$(\tau = N\xi,)$$

ξ = Column unity vector,
N = Fundamental matrix.

The fundamental matrix is defined as being

$$N = (I-Q)^{-1},$$

where I = Identity matrix,
Q = Submatrix of the transition matrix P.

Q is found by partitioning the transition matrix in the following manner:

$$P = \begin{bmatrix} I & \vdots & O \\ \hline R & \vdots & Q \end{bmatrix}$$

where O = Null matrix,
R = Matrix containing transition elements which describe the process entering an absorbing state,
Q = Matrix containing transient elements.

Application to Test Data

The mathematical model was applied to several prototype resistors to determine the possibility of detecting those elements which would fail before the end of design life. Based upon test data for three elements which extended for the length of time defined as design life, and two interval "acceptance" tests for subsequent elements, the predicted failure times for the latter were determined to be as follows:

Component	Predicted Failure, Time Intervals	Actual Failure, Time Intervals
1x64	> 20	> 20
1x65	> 20	> 20
1x66	> 20	> 20
1x67	15.0	14.1
1x68	6.5	2.9
1x69	19.3	> 20
1x611	15.0	14.6
1x612	> 20	> 20
1x613	15.0	13.9
1x615	> 20	> 20

Design life was considered 20 time intervals long for convenience, and represented the upper limit for accurate prediction. The acceptance tests were two time intervals in length.

Discussion of Results

The method used to detect in advance seemingly acceptable elements which failed prior to the end of design life was quite satisfactory. A rational approach for selecting the optimum range of performance for each performance level and the optimum length of each time interval remains to be developed.

It may be desirable to extend the method so that the basic transition matrix is modified to reflect the life history of each element as it becomes available (e.g., from telemetry data). Modification of the matrix to reflect the stress history of the element between the time of the acceptance test and the time when it enters service and any differences in service and test environments and conditions, appears worthwhile.

ASSUMPTIONS

The assumptions made for this study are as follows:

a. Component performance degrades continuously with passage of time. Smooth curves are fitted to the test data to obtain the transition matrix.

b. The component has no redundant elements. The use of redundant elements may be taken into consideration if necessary. The intent of this study was to determine initial failure, and redundancy was not a consideration.

c. The performance measurement is sensitive to the degradation of performance of the elements most likely to fail. Failures of elements whose degradation of performance cannot be detected by the measurement taken are considered chance failures. If multiple measurements are required, similar procedures to those described in this report would be followed.

d. Service environment is approximated in all important features during the extended tests used to obtain the transition matrix.

Any significant variation from the test conditions during actual service will yield failures occurring at times other than those predicted. The difference in predicted and actual failure would be a function of considerations which would include the severity of the added stresses and the sensitivity of the component to them.

ANALYSIS

General

In elementary theory of finite Markov processes, two basic types of chains (those with and those without transient sets) exist together with a multiplicity of modifications. The analysis can be greatly simplified by considering the performance characteristics of the equipment to be studied and particularizing the solution accordingly. This analysis is restricted to an element, module, or system whose performance is satisfactorily indicated by a single measurement. It is assumed the fact has been established that the measurement taken is sensitive to degradations of performance of the elements which will most probably fail, and failures of elements of the other category will be attributed to chance. Also, continuous degradation of performance with passage of time is assumed to exist. Once performance has reached a predetermined level defined as failure, the component is removed from service permanently. These last assumptions are made to simplify the analysis and to aid in applying the resultant model to test data. The assumptions should not be regarded as limitations of the theory. Since the performance of nonmaintainable electronic equipment seldom improves with the passing of time, and out-of-limits performance results in equipment functioning in some manner other than that intended, the assumptions pose no severe limitations.

The performance of the equipment to be analyzed is such that an absorbing Markov chain can be used. The properties and requirements of this chain will be reviewed and its use demonstrated.

Application of the Absorbing Chain to Performance Prediction

The finite Markov chain, as applied to electronic equipment performance, refers to a stochastic process which moves through a finite number of performance levels called states. This implies a consideration of time, and a division of the anticipated equipment output or measurement range into a number of operating levels, both of which will be considered in more detail later. The probability of entering any given level of performance is assumed to be a function only of the level occupied during the preceding time interval.

The term "absorbing," applied to the finite Markov chain, means that there exists at least one level of performance which,

when reached, cannot be left; i.e., when performance has dropped to some predetermined level (the absorbing state), the equipment will be considered to have failed. With only one level of performance defined as failure, the ergodic set consists of only one element, the absorbing state. Nonabsorbing levels, or states, of equipment performance are then termed transient, and in Markov theory a transient state is not reentered once it has been left.

The equipment performance considered here degrades with passage of time; also, the equipment is considered nonmaintainable. The nature of equipment performance degradation is compatible with the characteristics of the absorbing chain, and it can easily be proved that in any finite Markov chain, regardless of the level at which observations begin, as the number of observations approaches infinity, the probability that the process or performance has entered an ergodic state approaches unity.

Transition Matrix

Reference (1) provides a lucid definition of the Markov chain, which is readily adapted to equipment performance as follows:

The range of expected equipment performance is divided into a number of performance levels, termed outcomes (s_i). For a set of independent measurements, a set of possible outcomes, s_1, s_2, s_3 ... s_n, is possible.

A probability, p_k, is associated with each outcome. The multiplicative property, $p_r(s_{i0}, s_{i1}, s_{i2} \cdots s_{in}) = p_{i0}\ p_{i1} \cdots p_{in}$, defines the probabilities of sample sequences.

The result of any periodic performance measurement, which is recorded as one of a series of measurements during the testing of a component, is considered to depend only upon the outcome of the immediately preceding measurement. A fixed probability, p_k, associated with outcome (or level), s_j, is no longer adequate. It is appropriate now to speak of a conditional probability, p_{jk}, associated with every pair of outcomes (s_j, s_k). This is read as follows: given level of performance s_j as a result of some measurement during the component test, the probability that the level will be s_k at the next measurement is p_{jk}. If a_j is the probability that the first measurement will reveal the equipment to be at level s_j, then

$$P_r\ (s_j, s_k) = a_j p_{jk}. \tag{1}$$

In general, this can be written (1)

$$P_r\ (s_{j0}, s_{j1} \cdots s_{jn}) = a_{j0} p_{j0j1} p_{j1j2} \cdots \tag{2}$$

$$\cdots\ p_{j(n-2)j(n-1)} p_{j(n-1)j(n)}.$$

The conditional probabilities, p_{jk}, represent the probability of a transition from operating level s_j to level s_k, and are termed transition probabilities.

They may be written in matrix array as follows:

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \cdots \\ p_{21} & p_{22} & p_{23} \\ \vdots & & \end{bmatrix} \tag{3}$$

This forms a stochastic matrix.
Any stochastic matrix, together with the initial probability a_k, completely defines a Markov chain. (1).

Canonical Form of Transition Matrix

To permit orderly matrix manipulation of the transition matrix, P, the matrix is partitioned and arranged in the following manner:

> First, the states, or levels of performance, are numbered so that the elements of any given equivalance class are consecutively numbered. Two states are in the same equivalence class, for example, if they can communicate, or if one can go from either state to the other one. The sets are arranged so that minimal, or ergodic, sets are first, sets one level above minimal are next, and so on. For the problem at hand, this permits progress from one state to another in the same class, or to a state in a lower class, but not to a state

in a higher class. This corresponds to the expected performance trend, i.e., degradation.

Second, the transition matrix is arranged in a manner described by Kemeny and Snell (2).

$$P = \begin{array}{c} \begin{matrix} \overbrace{}^{r-q} & \overbrace{}^{q} \end{matrix} \\ \left[\begin{array}{c|c} S & O \\ \hline R & Q \end{array}\right] \end{array} \begin{matrix} \rbrace r-q \\ \rbrace q \end{matrix} \qquad (4)$$

where there are q transient states, and r ergodic states.

All ergodic sets are united to form S. In assuming minimum performance as a definition of failure, the S submatrix is but a single element.

The submatrix 0 is a null matrix.

The submatrix R consists of elements which deal with the transition from transient to ergodic states.

The submatrix Q consists of elements which are the transition probabilities for the process as long as it remains in transient states.

In considering an absorbing chain, the S matrix becomes an identity matrix, I.

As an example, a transition matrix will be formed from the performance data of three electronic elements denoted 1x61, 1x62, 1x63 (see Appendix).

$$P = \begin{array}{c|cccccc} & S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ \hline S_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline S_2 & 2/13 & 11/13 & 0 & 0 & 0 & 0 \\ S_3 & 0 & 2/17 & 15/17 & 0 & 0 & 0 \\ S_4 & 0 & 0 & 3/13 & 10/13 & 0 & 0 \\ S_5 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ S_6 & 0 & 0 & 0 & 0 & 1 & 0 \end{array} \begin{matrix} \rbrace -(r-q) \\ \rbrace -(q) \\ \\ \\ \\ \end{matrix} \begin{matrix} \\ \\ \\ \\ (r=6) \\ (q=5) \end{matrix} \qquad (5)$$

In this example, s_1 is absorbing and is written first. The regions I, O, R, and Q have been formed by the partitioning.

Fundamental Matrix

For an absorbing Markov chain, the fundamental matrix is defined (2) as

$$N = (I - Q)^{-1} \tag{6}$$

where, the superscript, -1, denotes the inverse matrix.

The question of whether (I - Q) possesses an inverse will be examined briefly.

The following identity is shown (2) by multiplying out the left side of the equation:

$$(I - Q)\ (I+Q+Q^2+ \ldots +Q^{n-1}) = I-Q^n \tag{7}$$

Regardless of where the process starts, in the finite absorbing Markov chain, the probability that the process is in an ergodic state (after n intervals) approaches unity as n approaches infinity. It can be seen, then, that Q in Eq. (4) raised to high powers, approaches zero, and the last q columns of P are zero.

It can be hypothesized that Q^n approaches zero as n approaches infinity. The right side of Eq. (7) approaches I, whose determinant is unity. For high enough values of n, $I-Q^n$ must, therefore, have a nonzero determinant. From rules of matrix manipulation it is known that the determinant of the product of two matrices is equivalent to the product of the determinants of the two matrices; hence, I-Q cannot have a zero determinant, which is a sufficient condition for (I-Q) to have an inverse. Since (I-Q) can have an inverse, $(I-Q)^{-1}$, both sides of Eq. (7) can be premultiplied by it, yielding

$$(I + Q + Q^2 + \ldots + Q^{n-1}) = (I - Q)^{-1}\ (I - Q^n) \tag{8}$$

The right side of Eq. (8) goes to $(I - Q)^{-1}$ from the preceding; and since it has been shown that the inverse of (I - Q) can exist, it becomes

$$(I - Q)^{-1} = I + Q + Q^2 \ldots + \sum_{k=0}^{\infty} Q^k \qquad (9)$$

Expected Value and Variance

The following example illustrates the nomenclature used.

A component, A or B, is selected at random for testing over a period of two time intervals. Performance measurements will be taken at the end of each time interval. The following possibilities are known to exist:

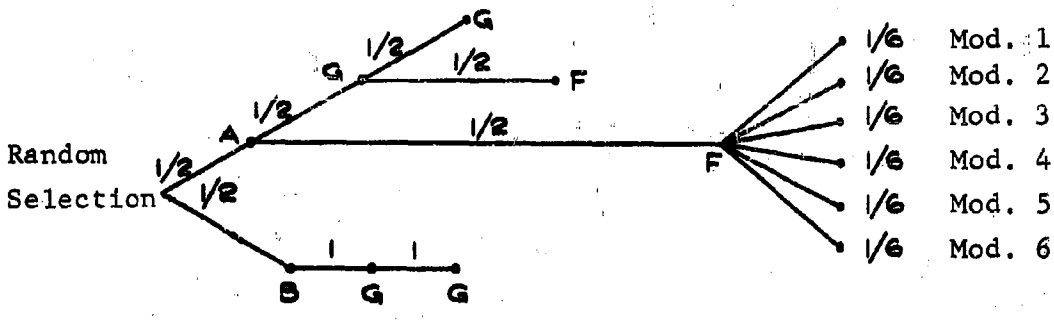


Figure 1. Paths and Tree of Component Randomly Selected for Testing Over Two Time Intervals

In this type of representation, the sequence of events leading to any conclusion is termed a path, and the array of paths, a tree. Each path comprises two branches of the tree; conditional probabilities of proceeding from one state to a succeeding state are shown on each branch.

Satisfactory performance is denoted by G, and failure by F. Then, having selected A, the probability that the equipment will have failed by the end of the first interval, and the probability that the cause of the failure will be attributed to module No. 1 is (1/2) (1/6) = 1/12. These known probabilities are called branch probabilities, and the product of the branch probabilities leading to any conclusion is called the path weight. Note that the sum of the weights of all the possible paths equals unity.

The set of all possible paths is called a possibility space, and is designated by U.

The same general procedure as that shown in Fig. 1 can be used for any test in which the measurements are recorded at regular intervals. The only requirement is that there be a finite number of outcomes possible at each stage and that the probabilities for any given outcome at any stage, given the outcomes at the preceding stages, be known.

From Fig. 1, the following table can be prepared:

TABLE 1. POSSIBILITIES, WEIGHTS AND OUTCOME FUNCTIONS FOR TWO RANDOMLY SELECTED COMPONENTS

t	w(t)	$f_1(t)$	$f_2(t)$	$f_3(t)$
1	1/8	A	G	G
2	1/8	A	G	F
3	1/24	A	F	1
4	1/24	A	F	2
5	1/24	A	F	3
6	1/24	A	F	4
7	1/24	A	F	5
8	1/24	A	F	6
9	1/2	B	G	G

where t = possibilities
w = weight
f = outcome functions.

Let f equal a function with domain U and range T, and assume that a measure has been assigned to U.

For each t in T, let

$$w(t) = Pr\ (f=t) \tag{10}$$

Then the weights w (t) give the probability measure on the set T, which will be called the induced measure for f. The weights will be called induced weights, and the induced measure can be illustrated, using both range and weights, as follows:

$$f: \begin{bmatrix} t_1 & t_2 & t_n \\ w(t_1) & w(t_2) & (wt_n) \end{bmatrix} \tag{11}$$

For the example, let f be the function which gives the number of times that A is shown operating properly, independent of B. The range of f is

$$T = (0, 1, 2) \tag{12}$$

Then, Pr (f = 0) = 1/2; Pr (f=1) = 1/4; Pr (f=2) = 1/4. The range and induced measure, then, are

$$f: \begin{bmatrix} 0 & 1 & 2 \\ 1/2 & 1/4 & 1/4 \end{bmatrix} \tag{13}$$

When f is defined on the possibility space

$$T = (t_1, t_2, t_3 \ldots t_n) \tag{14}$$

and assuming that a measure, w (t_i), has been assigned, the mean or expected value of f, M(f), is

$$M(f) = \sum_i f(t_i) \cdot w(t_i) = \sum_i t_i \cdot w(t_i) \tag{15}$$

From Eq. (13), for the example,

$$M(f) = 0 \cdot 1/2 + 1 \cdot 1/4 + 2 \cdot 1/4 = 3/4 \tag{16}$$

Letting M (f) = m, the variance of f, Var (f), is the second moment about the mean or the mean of the function $(f - m)^2$.

Substituting into Eq. (15) yields

$$M(f-m)^2 = \sum_j (f_j^2 - 2mf_j + m^2)\, g(f_j) \tag{17}$$

Multiplying out the right side, and substituting from above yields

$$M(f^2) - 2\,m\,M(f) + m^2 = M(f^2) - m^2 \quad (18)$$

$$= \text{Var}(f)$$

Application of Expected Value to Performance Prediction

To predict the time of operation required before an equipment enters the absorbing state or fails, Let f equal the function which gives the number of time intervals that the process is in a transient state (i.e., the number of time intervals, including the first, that the equipment operates satisfactorily).

If the process begins in an ergodic state, f = 0. If the process begins in a transient state, f is the number of time intervals required for the process to reach the erogodic state; i.e., it is the time required to reach absorption.

Let n_j be the function which gives the number of time intervals that the performance is in state (or level) s_j. The function n_j is defined only for the transient states, s_j. Let δ^k_j be a function that is unity if the performance is in state s_j after k time intervals and is zero otherwise.

It can be seen that

$$n_j = \sum_{k=o}^{\infty} \delta^k_j \quad (19)$$

Then, as shown by Kemeny and Snell (2),

$$M_i(n_j) = M_i\left(\sum_{k=o}^{\infty} \delta^k_j\right) \quad (20)$$

$$= \sum_{k=o}^{\infty} M_i(\delta^k_j)$$

$$= \sum_{k=o}^{\infty} \left((1-p^{(k)}_{ij}) \cdot 0 + p^{(k)}_{ij} \cdot 1\right)$$

$$= \sum_{k=o}^{\infty} p^{(k)}_{ij} = \sum_{k=o}^{\infty} Q^k \text{ (since } s_i, s_j \text{ are transient)}$$

$$= N, \text{ from Eq. (6) and Eq. (9)}$$

From the definition of f,

$$f = \sum_{s_j \in T} n_j \quad (s_j \text{ are the transient states in the set T}) \tag{21}$$

Then,

$$M_i(f) = \sum_{s_j \in T} M_i(n_j) \tag{22}$$

(From Eq. (19) and Eq. (20.)

$$M_i(f) = \sum_{s_j \in T} N \tag{23}$$

$$= N\xi$$

where ξ = unity column vector to give the row sums of N. Each row of N corresponds to one of the transient states, s_j.

The variance of f,

$$\mathrm{Var}_i(f) = M_i(f^2) - M_i(f)^2 \quad . \quad \text{From (2)}, \tag{24}$$

$$M_i(f^2) = (2N-I)(N\xi) \tag{25}$$

From Eq. 23,

$$M_i(f)^2 = (N\xi)^2 \tag{26}$$

Substitution into Eq. 24 then yields

$$\mathrm{Var}_i(f) = (2N - I) N\xi - (N\xi)^2 \tag{27}$$

DEMONSTRATION OF METHOD

Method of Analysis

In the missile and space vehicle industry, conditions often preclude extensive life testing. The situation investigated here assumes that it is possible to test three components of a model under development for the length of time that the component is expected to operate in service, including testing (design life). During the testing of the three components, simulated service loads and environments were maintained.

The period of time representing design life was divided into 20 time intervals. The choices of time-interval length and range of operating performance, from a maximum expected level to a level termed failure, were divided into a convenient number of performance levels. A transition matrix was then formed for three prototype resistors as illustrated in the Appendix, and given by Eq. (28) to demonstrate the method. The matrix could have been established analytically and modified as data became available.

Eleven additional elements were made available for similar environmental tests to be conducted over two time intervals. The object of these tests was to determine if elements whose performances would degrade to failure before the twentieth time interval, after they were placed in service, could be identified on the basis of observed performances during the abbreviated tests.

The sequence of events was as follows:

a. Performance measurements for resistors 1x64 through 1x613 and 1x615 were recorded for two time intervals.

b. The transition matrix was modified to reflect the change of performance level for each element to utilize significant a posteriori knowledge.

c. The mean-time-to-failure was predicted for each on the basis of the modified transition matrix and the level of performance existing at the end of the second time interval.

If telemetry data were available, or if the performance measurement could be continued in service until the end of design life or until each component failed in service, the original transition matrix could be modified to utilize this additional knowledge, if necessary.

Curve Fit

Life-test data for a number of selected electronic elements was obtained to demonstrate the results of the preceding section. Qualitative performance of each element was indicated by a single measurement recorded as deviation from component nominal-rated performance. The IBM program,

"Multiple Linear and Nonlinear Regression Analysis for the IBM 1620", was used to obtain performance trend curves from the data. A fourth-degree polynominial fit was used for each set of data. The resulting curves plotted for the period of time designated as design life appear as Figures 2, 3, and 4.

Transition Matrix

The development of a transition matrix using elements 1x61, 1x62, and 1x63 is given the Appendix; the matrix is shown below:

$$P = \begin{array}{c|cccccc} & S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ \hline S_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ S_2 & 2/13 & 11/13 & 0 & 0 & 0 & 0 \\ S_3 & 0 & 2/17 & 15/17 & 0 & 0 & 0 \\ S_4 & 0 & 0 & 3/13 & 10/13 & 0 & 0 \\ S_5 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ S_6 & 0 & 0 & 0 & 0 & 1 & 0 \end{array} \qquad (28)$$

Expected mean times until failure can be predicted for future components of the same model on the basis of a short test, relative to design life, using the transition matrix. The object of such predictions is to identify components whose performances are degrading in such a manner that they would fail before expiration of the design life when operating under the same conditions in service.

Failure is arbitrarily set as -0.25% deviation from nominal; i.e., from the components under study, the designer requires performance of a level that does not equal or drop below -0.25% of nominal value during design life when the components are operating under identical load and environment provided for life-test components.

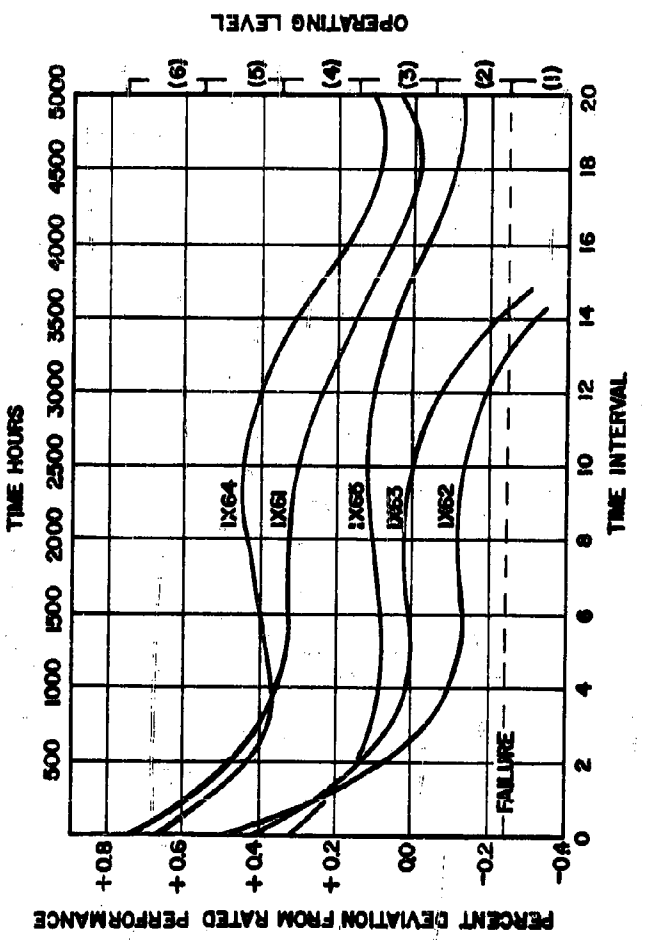


Fig. 2 Performance Curves for Components 1x61 Through 1x65

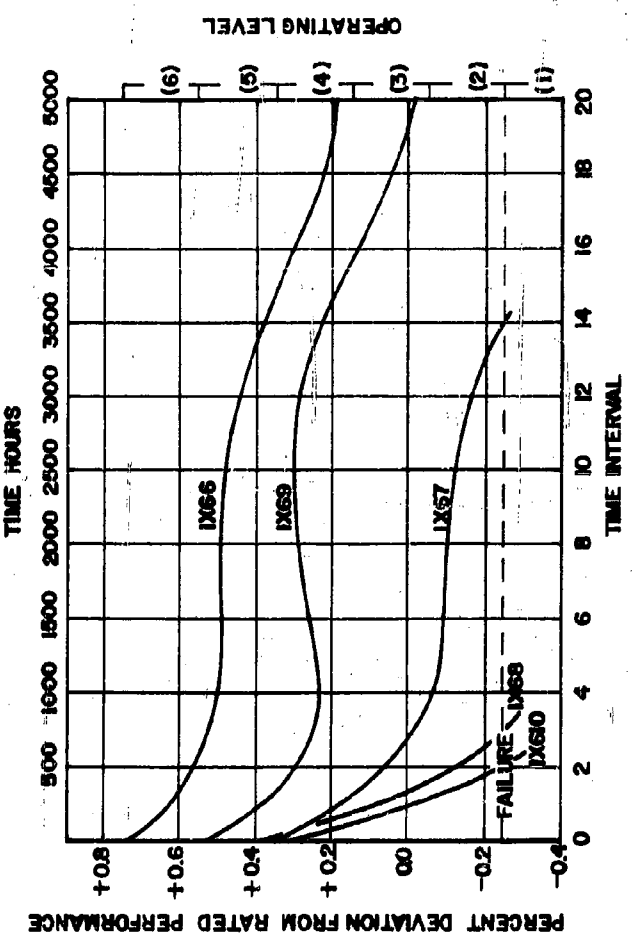


Fig. 3 Performance Curves for Components 1x66 Through 1x69, and 1x610

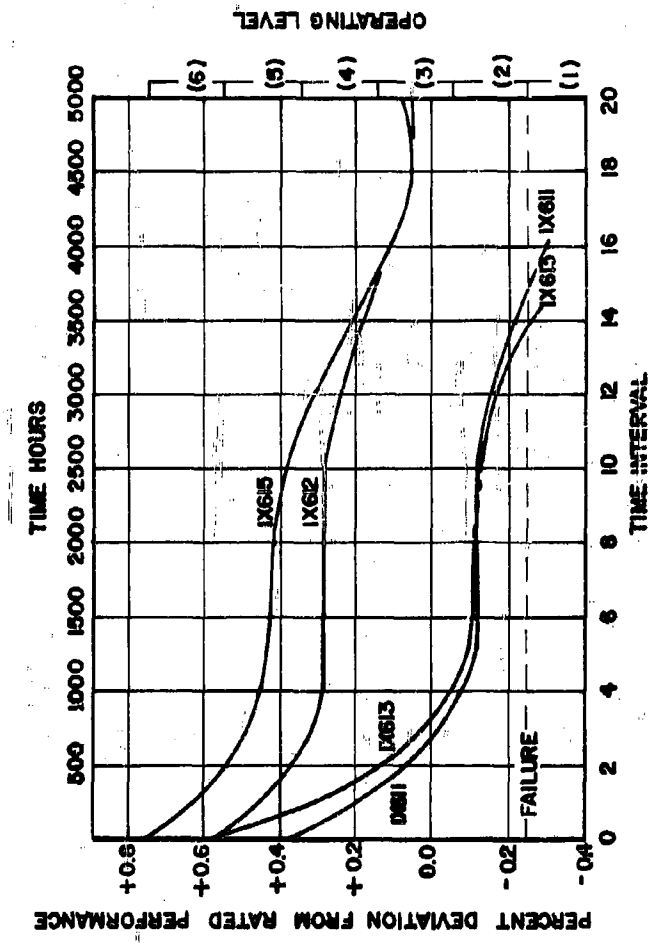


Fig. 4 Performance Curves for Components 1x611, 1x612, 1x613, and 1x615

Fundamental Matrix

The transition matrix, Eq. (28), is partitioned in accordance with Eq. (4) as follows:

$$P = \begin{bmatrix} I & 0 \\ R & Q \end{bmatrix} = \left[\begin{array}{c|ccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 2/13 & 11/13 & & & & 0 \\ 0 & 2/17 & 15/17 & 0 & 0 & 0 \\ 0 & 0 & 3/13 & 10/13 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right] \qquad (29)$$

On the basis of a two-interval test (see Figures 2, 3 and 4), the expected times to failure of components 1x64 through 1x69, 1x611 through 1x613 and 1x615 will be calculated using Eq. (29). During the second interval, 1x610 failed and incorrect computer input was used for 1x614, invalidating its performance curve.

Calculations for 1x64:

$$P_4 = \left[\begin{array}{c|ccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 2/13 & 11/13 & 0 & 0 & 0 & 0 \\ 0 & 2/17 & 15/17 & 0 & 0 & 0 \\ 0 & 0 & 3/13 & 10/13 & 0 & 0 \\ 0 & 0 & 0 & 2/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right] = \begin{bmatrix} I & 0 \\ R & Q \end{bmatrix} \qquad (30)$$

The above matrix, Eq. (30) is obtained by modifying Eq. (29) in accordance with the test results.

Fundamental Matrix:

$$N = (I-Q)^{-1} = \frac{\text{Adj. } (I-Q)}{\det. \ (I-Q)} \qquad (31)$$

$$N = \begin{bmatrix} 13/2 & 0 & 0 & 0 & 0 \\ 13/2 & 17/2 & 0 & 0 & 0 \\ 13/2 & 17/2 & 13/3 & 0 & 0 \\ 13/2 & 17/2 & 13/3 & 3/2 & 0 \\ 13/2 & 17/2 & 13/3 & 3/2 & 1 \end{bmatrix} \qquad (32)$$

and from Eq. (23),

$$N\xi = \tau = \begin{matrix} S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{matrix} \begin{bmatrix} 13/2 \\ 15 \\ 58/3 \\ 125/6 \\ 131/6 \end{bmatrix}$$

At the end of the second interval, 1x64 was in level 5. From Eq. (32) τ_5 = 125/6 time intervals, predicted time until failure.

(33)

Since the transition matrix was formed from data which extended only 20 time intervals, the conclusion which can be drawn is that the element probably will not fail during the 20 interval period defined as design life.

A summary of the predicted and actual failures is given below for the remaining elements:

TABLE 2. SUMMARY OF PREDICTED AND ACTUAL FAILURES FOR REMAINING COMPONENTS

Element	Predicted Failure, Time Intervals	Actual Failure, Time Intervals
1x64	> 20	> 20
1x65	> 20	> 20
1x66	> 20	> 20
1x67	15.0	14.1
1x68	6.5	2.9
1x69	19.3	> 20
1x611	15.0	14.6
1x612	> 20	> 20
1x613	15.0	13.9
1x615	> 20	> 20

APPENDIX

TIME INTERVALS AND RANGES OF OPERATING LEVELS

There are a number of ways to divide the performance curves into convenient time intervals and ranges of operating levels; the method presented here is one method, but not the only method. The time intervals and operating levels picked for components 1x61, 1x62 and 1x63 were chosen because the smallest number of levels was desired (consistent with acceptable sensitivity to level changes) to hold the number of calculations to a minimum. The intervals were adjusted so that, in general, changes in more than one level of performance were not likely to occur during one time interval.

Use of computer facilities and possible improvement in accuracy with optimization of both length-of-time interval and range of performance within one operating level should be considered when choosing these parameters.

The assignment of "operating levels" is for convenience, and is not from necessity. Actual performance measurements (watts, dimensionless ratios, etc.) may be used in place of operating levels if desired.

APPENDIX

The following table was prepared from Fig. 2 (3)

Part No.	State at the End of Each Interval																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1x61	6	5	5	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3
1x62	4	3	3	2	2	2	2	2	2	2	2	2	2	1						
1x63	4	3	3	3	3	3	3	3	3	3	3	2	2	2	1					

Frequency of Passage from State S_i to S_j in One Interval

S_{65}	S_{55}	S_{54}	S_{44}	S_{43}	S_{33}	S_{32}	S_{22}	S_{21}	
1	1	1	10	1	5				
				1	1	1	9	1	
				1	9	1	2	1	
1	1	1	10	3	15	2	11	2	TOTAL

APPENDIX

Tabulating the frequencies in an array,

	S_1	S_2	S_3	S_4	S_5	S_6	Σ
S_2	2	11	0	0	0	0	13
S_3	0	2	15	0	0	0	17
S_4	0	0	3	10	0	0	13
S_5	0	0	0	1	1	0	2
S_6	0	0	0	0	1	0	1

$P =$

	S_1	S_2	S_3	S_4	S_5	S_6
S_1	1	0	0	0	0	0
S_2	2/13	11/13	0	0	0	0
S_3	0	2/17	15/17	0	0	0
S_4	0	0	3/13	10/13	0	0
S_5	0	0	0	1/2	1/2	0
S_6	0	0	0	0	1	0

Dividing each element by its row sum yields, P, the transition matrix.

NOMENCLATURE

s	Performance level
p	Probability
P	Transition matrix
q, r	Number of transient and ergodic states
S	Matrix composed of ergodic sets
I	Identity matrix
O	Null matrix
R	Matrix of elements dealing with transition from transient to ergodic states
Q	Matrix of transition probabilities for process as long as it remains in transient states
N	Fundamental matrix
n, k	Integers
U	Possibility space
t	Possibility
w	Weight
f	Outcome function
T	Range of outcome function
M	Mean or expected value
$\mathbf{e}$	Column unity vector
δ_j^k	Function that is unity of performance is in state sj after k time intervals and zero otherwise
τ	Expected failure time

REFERENCES

1. Feller, W., An Introduction to Probability Theory and Its Applications, John Wiley and Sons, Inc., New York, 1950.

2. Kemeny, J. G., and J. L. Snell, Finite Markov Chains, D. Van Nostrand Company, Inc., Princeton, N. J., 1960.

3. Bradley, C. E., and E. L. Welker, "A Model for Scheduling Maintenance Utilizing Measures of Equipment Performance", ARINC Research, Monograph No. 8 (Pub. No. 101-21-150), October, 1959.

RELIABILITY AS A THERMOSTRUCTURAL DESIGN CRITERION

F. W. Diederich
W. C. Broding
A. J. Hanawalt
R. Sirull

Avco Corporation
Research and Advanced Development Division
Wilmington, Massachusetts

ABSTRACT

An approach is outlined for using reliability as a design criterion for thermostructural and similar systems in lieu of sets of safety factors for combinations of various environments; the primary objective of which is to avoid the sometimes large, but almost invariable unknown, degree of conservatism introduced by the combination of safety factors.

Several methods are indicated for implementing this approach with emphasis on a small-perturbation method and a Monte-Carlo method. The estimation of confidence limits for the calculated reliability in terms of an implied sample size determined by the uncertainty of the available information is discussed.

The use of this approach in failure-mode analyses, reliability apportionment, and similar reliability calculations is indicated briefly.

The work described in this paper was performed under Air Force Contract Nos. AF04(647)-16 and -36.

RELIABILITY AS A THERMOSTRUCTURAL DESIGN CRITERION

F. W. Diederich, W. C. Broding, A. J. Hanawalt, R. Sirull
Avco Corporation
Research and Advanced Development Division
Wilmington, Massachusetts

INTRODUCTION

The design of a system or component to withstand an environment, or combination of environments, invariably involves some uncertainty because environments and the performance of the system or component under them can be defined only within a certain precision. Consequently, the traditional design practice is to incorporate a measure of conservatism to assure satisfactory performance under all, or almost all, conditions.

In structural, thermal, and similar designs, this conservatism is usually specified by a safety factor, the ratio of the allowable stress, temperature, etc., to the expected value (i.e., the nominal, average, or "design" value). This safety factor is intended to make up for:

1. Uncertainties in the environment or the performance under it; presumably, as a result of inability to define it beyond certain limits.

2. Randomness in the environment or the performance under it; the statistical variation from case to case as a result of differences in the applied load, heating, etc., and in the various properties such as thicknesses (all within specified limits on tolerances).

3. Inability to calculate or measure the performance under the proper type of environment; for instance, making static calculations when dynamic conditions are critical, and making separate performance calculations or tests for several environments when a combined environment is critical.

Clearly, excessive safety factors imply unnecessary weight and cost, and where these considerations are important, a great incentive exists for reducing safety factors to the minimum required value. The traditional approach is to make more and more sophisticated tests and calculations to eliminate any unnecessary conservatism introduced by the third reason listed above. This leaves the other two, however.

In relatively simple problems, such as a single critical environment, experience can often serve as a guide in arriving at a safety factor which furnishes the required assurance against failure without excessive weight penalty. Even here however, experience can be misleading. For instance, a safety factor of 1.25, for structural failure of a metallic component may be entirely adequate, whereas the same factor for a plastic or ceramic element (with greater variation in the mechanical properties) might be inadequate, even if both are based on "conservative" properties.

This example indicates that to fulfill its functions, the safety factor must be related to the statistical variation in the performance under the given environment (and in the environment itself). This suggests the specification of a safety factor on the basis of statistical consideration, the desired reliability (probability of survival under the given environment). See Figure 1.

Suppose, for instance, that an applied load is subject to a random variation (with a Gaussian distribution) of 5 per cent (1σ). A design for safety factor of 1.2 based on the average load then implies that failure represents a 4σ condition, so that the reliability is 0.9997.

Similarly, if the load is invariant within very narrow limits but the allowable strength (or modulus, in the case of buckling failure) is subject to a random (Gaussian) fluctuation of, say, 5 per cent (1σ), a safety factor of 1.2 again implies a reliability of 0.99997.

Lastly, consider the case where the load is subject to a 4 per cent (1σ) variation and the strength (or modulus) to a 3 per cent variation. Now, one can consider the safety factor in any stated case (given combination of load and strength) to be itself a random process with characteristics defined by the statistics for the load and strength, respectively (see Appendix A). Thus, if the load and strength are statistically independent, the standard deviation in the safety factor is approximately 5 per cent (obtained by root-mean-squaring 3 and 4 per cent), a procedure which is now valid only as an approximation for standard deviations that are small compared with the mean value (see Appendix A) so that the failure again represents a 4 to 5σ case.

This approach of tieing in the safety factor with reliability considerations places the effect of randomness (item 2 above) on a rational basis. It remains to take the effect of uncertainty into account in a similar manner. Uncertainties, in this context, stem from several sources, but are basically of two kinds. First, the materials properties are subject to uncertainty due to the finite (often very small) size of the sample (number of test articles) that was used to define them. The same situation prevails in connection with the environment (say, atmospheric density at high altitudes, which influences the heating environment). These uncertainties are usually expressed in terms of confidence limits. Second, in the cases of some sophisticated properties (transient thermodynamic properties for instance) and environments,

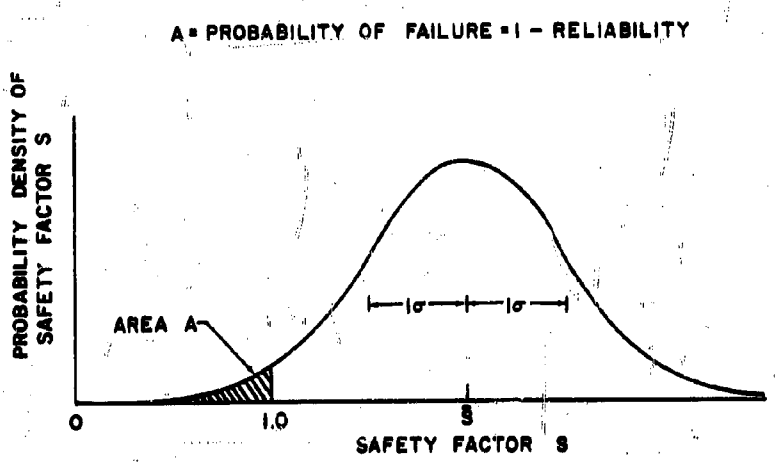


Fig. 1. Probability Density of Safety Factor

there are limitations on instrumentation and techniques which would, even with infinite sample size, leave some doubt as to the precise properties. The instrument used, for instance, may be good only to ± 3 per cent; this figure being arrived at by an analysis of the workings of the instrument, calibration (possibly under simplified conditions) against a more precise one, or by an educated guess, as the case may be. The implication then is that the "true" value could be anywhere within a 3 per cent bracket, with no way of differentiating just where it is. It is a fallacy, then, to assume that the probability of its being at any given place is uniform, or Gaussian, or anything else; this is not a random process, and probability is a meaningless term when applied to it. The whole subject of uncertainty and its relation, if any, to randomness is a very deep one and will not be gone into here. For the purposes of the present analysis, it is possible to relate such uncertainties to confidence limits, and thus, interpret them in a way which is generally understood (see Appendix B).

The basic philosophy under discussion, thus, consists in using a statistical understanding of the randomness of the problem to relate a safety factor to reliability, and a knowledge of the uncertainties to furnish an indication of the confidence attached to the reliability calculation. The approach can be used, therefore, either for calculating the reliability of a given design or for effecting a design to a given reliability criterion.

For the simple examples mentioned previously, the proposed approach, although useful, does not generally represent a really significant improvement over the more traditional philosophy. The situation where the reliability-criterion approach becomes very much preferable on many counts is the one where the design has to be based on a complicated combination of environments, acting both sequentially and simultaneously.

One such example is the thermostructural performance of a re-entry vehicle. Typical re-entry conditions are shown in Figure 2. Here, the thermal environment (heating input) results in a response of the system (ablation and temperature rise of the heat-shield/structure composite). This gives rise to changed material properties, and to induced thermal stresses, (in several possible modes, depending on the performance of the bond, etc.), which interacts with the stresses induced by the longitudinal and lateral loads.

The use of "conservative" properties, safety factors on stresses and heating, in conjunction with a detailed thermostructural analysis (based on conservative assumptions) and with a thorough test program to uncover any unforeseen flaws, will generally result in a safe design, but does not furnish any quantitative insight into the degree of overall conservatism implied by this approach. Some studies have shown that even with very low safety factors (say, 1.25 on loads, 1.15 on heating), one is likely to arrive at

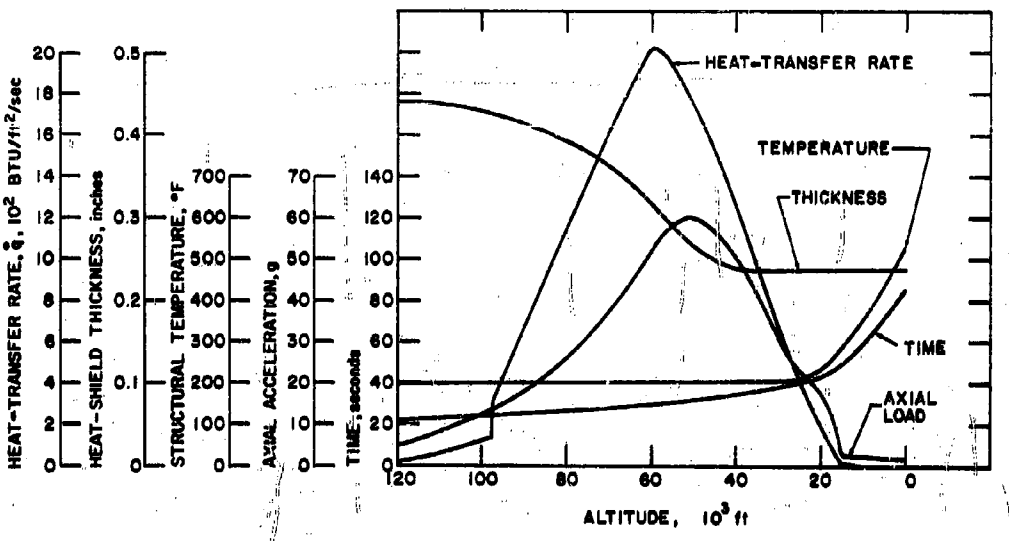


Fig. 2. Typical Re-entry Conditions

unnecessarily conservative designs. For this type of problem, the application of a reliability criterion results in an approach which, in conjunction with a correct analysis of the failure modes, furnishes a design with a known degree of conservatism; that is, a known reliability (with a given confidence).

The purpose herein is to outline this approach. Most of the emphasis will be placed on its application to the aforementioned thermostructural problem, both from the design and reliability-prediction points of view.

SYMBOLS

a	Plate length
b	Plate width
C_p	Specific heat
E	Young's modulus
F	Allowable stress
f	Actual stress
fn	Notation for "a function of"
H_v	Heat of vaporization
K	Plate-buckling coefficient
k	Conductivity
N	Sample size
$\hat{N}$	Implied sample size
P	Load
p	Parameter
Q	Cold-wall heat input
Q^*	Effective heat of ablation

q	Heat-transfer rate
R	Cylinder radius
r	Reliability (probability of survival)
S	Safety factor
s	Normalized safety factor, $(\bar{S}-1)/\sigma_s$
T	Temperature
t	Thickness
V	Re-entry speed
W	Weight
α	Coefficient of thermal expansion
γ	Re-entry angle
ϵ	Uncertainty (as a fraction of the mean value)
η	Blowing factor, also emissivity in Table 1
ν	Poisson's ratio
π_i	Internal pressure
π_o	External pressure
ρ	Density (of heat shield, unless otherwise specified)
σ	Standard deviation
τ	Time
Φ	Normalized Gaussian distribution function
ϕ	Normalized Gaussian probability density
ψ	Correlation coefficient

Subscripts

a	Ablated condition
i	i^{th} parameter

L Laminar

S Safety factor

s Structure

T Turbulent

Superscript

* Reference condition (except on Q, see above)

STATISTICAL BASIS

The basic principle can be illustrated with a very simple example; namely, a structural element loaded in tension. Suppose the load and the size of the member are very closely controlled, but the strength is variable from specimen to specimen about a mean allowable stress $\overline{F}$; suppose, furthermore, that the probability distribution of the allowable is Gaussian, with standard deviation σ_F. The actual safety factor,

$$S = \frac{F}{f} \tag{1}$$

also varies from specimen to specimen with a Gaussian distribution, with mean $\overline{S} = \overline{F}/\overline{f}$, standard deviation σ_S (see Figure 1 and Appendix A). The cases that fall below $S = 1$ represent failure, and those for which $S > 1$ represent success or survival. The fraction of the cases for which $S > 1$ is, therefore, the probability of survival; i. e., the reliability; and this fraction can be obtained from a table of "areas under the normal curve"; that is, of the Φ function (see Appendix A) for the given normalized safety factor,

$$s = \frac{\overline{S} - 1}{\sigma_S} \tag{2}$$

The simplicity of the example is due to the presence of only one variable; namely strength. However, in general, the load, the element size, and the strength all vary randomly as a result of random deviations in the applied heating (which, in turn, depends on a number of somewhat random factors), the elastic properties of the element at elevated temperatures, etc. Consequently, the safety factor may be considered to be a function of several parameters p_i pertaining to the applied environments and to the resistance to these environments (representing normalized aerodynamic cold-wall heating, atmospheric density, trajectory deviations giving rise to heating changes, several thermal properties, and ablation characteristics, several kinds of mechanical or thermomechanical properties for instance). This function will, in general, represent a complicated nonlinear dependence on the various parameters, which are, in turn interrelated in a complex nonlinear manner. The problem is to deduce the statistical characteristics of S from

those of the several parameters. These are assumed to have Gaussian distributions, in general, although other distributions are equally admissible. They are also assumed to be statistically independent, although the effect of correlation of parameters will be discussed.

There are three approaches to this problem:

1. Convolution (direct synthesis)

2. Monte-Carlo methods

3. Small-perturbation methods.

The first consists in the direct calculation of the desired probability density by an integration process involving the individual probability densities of the parameters and the relation which expresses S in terms of the parameters (see Appendix A). Except in the simplest cases when this relation can be expressed in simple analytical form, so that the integration can be performed explicitly, this approach is not too practical. It requires not only the numerical definition of the relation between S and the parameters for very many combinations of parameters, but also complicated numerical integration of the results (with subtle questions involving truncation errors, etc.). The computing effort increases exponentially with the number of parameters.

The Monte-Carlo approach consists in generating artificially a statistical sample of the parameters having the proper separate (and, if appropriate, joint) frequency functions, calculating the safety factor for each case in the sample, and thus, generating a sample of safety factors from which its statistical characteristics may be deduced. In this approach, the number of calculations is independent of the number of parameters and depends essentially only on the desired confidence in the results, which is usually limited by the certainty to which the required input information is known. For highly nonlinear dependence of the safety factor on the input parameters, and to some extent, for heavily non-Gaussian probability densities of the parameters, the probability density of the safety factor may be significantly non-Gaussian, and thus, require a greater number of calculations to define the desired characteristics.

The third approach is based on the premise that, in many cases, the standard deviations of the input parameters are small (compared with the mean values), so that all of the input parameters (and, consequently, the corresponding safety factors) will cluster within a small domain of the average case. In this domain, the functional dependence of the safety

factor on the input parameters should be represented with sufficient accuracy by the linear terms of a Taylor series expansion; i.e.,

$$S = S^* + \sum_{i=1}^{n} \left(\frac{\partial S}{\partial p_i}\right)^* (p_i - p_i^*) \tag{3}$$

The assumption implied is that in the domain of interest (say, $\pm 3\sigma$ variations in the parameters), the influence of higher-order terms is negligible. In many cases of interest, this is true; but occasionally, heavy nonlinearities, sometimes approaching the nature of discontinuities, occur in the function relating the safety factor to the parameters within the domain of interest, in which cases the small-perturbation approach cannot be used.

The mean and standard deviation of S can be calculated directly from Eq. (3) and are

$$\bar{S} = S^* + \sum_{i=1}^{n} \left(\frac{\partial S}{\partial p_i}\right)^* (\bar{p}_i - p_i^*) \tag{4}$$

and

$$\sigma_S^2 = \sum_{i=1}^{n} \left(\frac{\partial S}{\partial p_i}\right)^{*2} \sigma_i^2 \tag{5}$$

if all parameters are statistically independent. If some parameters are dependent, additional terms arise on the right-hand side of Eq. (5); for instance, for a correlated pair p_r and p_s, a term,

$$2 \left(\frac{\partial S}{\partial p_r}\right)^* \left(\frac{\partial S}{\partial p_s}\right)^* \sigma_r \sigma_s \psi_{rs}, \quad m \tag{5a}$$

has to be added, where ψ_{rs} is the correlation coefficient for the two parameters.

This last approach is by far the simplest of the three to use; it requires only as many calculations of the safety factor as there are parameters to establish the partial derivatives (by numerical differentiation); the subsequent arithmetic requires very little effort. The derivatives are themselves useful byproducts because they furnish an insight as to which parameters are really significant in affecting the safety factor (and, hence, the reliability). With this insight, it is possible to improve the design by improving the critical characteristics either by modifying the design and/or environment in the areas of the critical parameters, or by establishing

closer specifications and holding closer tolerances on the critical parameters themselves and/or on factors affecting them.

Consequently, it is desirable to use this approach even in cases where heavy nonlinearities preclude its use in direct fashion. There are two ways in which the approach can be used in the presence of nonlinearities.

One way is to minimize the effect of nonlinearities by a judicious choice of the reference condition; an optimum choice of a reference condition is as follows. Of all possible combinations of input parameters, there is a subclass of cases which leads to failure; that is, for which $S<1$. The average condition of this subclass is a useful reference because Eq. (3) should then usefully describe the dependence of S on the parameters for cases which lead to failure, and, hence, to permit an accurate calculation of the fraction of cases which fail and, thus (by subtracting this fraction from 1) of the reliability. Unfortunately, the accurate location of the dividing surface and of the average of all conditions leading to failure requires lengthy calculations. However, for many purposes, an approximate average can be estimated from the physics of the problem. If the adequacy of the estimate is in question, another condition can be chosen and the sensitivity of the reliability calculation to the choice of reference condition established. If the sensitivity is great, another approach must be used.

Another way of handling nonlinearities is a combination approach, in which the nonlinear aspects are handled separately by a Monte-Carlo technique, and the results are combined with those for the linear aspects into a small-perturbation analysis. Both of these approaches have to be tailored to the needs of individual problems. Their use will be illustrated in the following sections.

THE LINEARIZED-PERTURBATION APPROACH

A Structural Example (Elastic Buckling of a Plate)

Consider a flat plate buckling under compressive end loads. The allowable failure stress is given by

$$F = \frac{KE}{1-\nu^2}\left(\frac{t}{b}\right)^2 \tag{6}$$

where the coefficient K is a function of the plate aspect ratio a/b. The applied stress is

$$f = \frac{P}{bt} \tag{7}$$

and the safety factor is

$$S = \frac{F}{f} = \frac{KE\,t^3}{(1-\nu^2)\,bP} \tag{8}$$

Suppose now that the applied load P, the thickness t, and the modulus E are subject to random variations by small amounts about their nominal values, whereas the other parameters are essentially fixed at their nominal values.

The linearized version of Eq. (3) is, then,

$$S = S^* + \left(\frac{\partial S}{\partial t}\right)^* (t - t^*) + \left(\frac{\partial S}{\partial E}\right)^* (E - E^*) + \left(\frac{\partial S}{\partial P}\right)^* (P - P^*) \tag{9}$$

where the derivatives can for this case be obtained analytically; namely,

$$\left(\frac{\partial S}{\partial t}\right)^* = \frac{3}{t^*}\,S^* \tag{10}$$

$$\left(\frac{\partial S}{\partial E}\right)^* = \frac{1}{E^*}\,S^* \tag{11}$$

and

$$\left(\frac{\partial S}{\partial P}\right)^* = -\frac{1}{P^*}\,S^* \tag{12}$$

The mean value and standard deviation of the safety factor can, then, be obtained from Eqs. (4) and (5) with these derivatives assuming that t, E, and P are statistically independent.

For this nonlinear problem, it is important to choose a reference condition of E, P, and t (and, hence, S), such that the nonlinearities are minimized. Equation (8) serves to establish combinations of t, E, and P for which $S<1$, and an average condition can then be estimated.

The reliability can then be found from the normalized safety factor s for any combination of E^*, t^*, and P^*, or for that matter, as a function of these parameters, thus permitting the selection of a thickness (and/or material), which, for a given nominal load, yields the desired reliability.

It may be remarked that even for this simple case, the convolution approach cannot be used, as indicated in Appendix A, because with a Gaussian distribution the possibility of having zero or negative loads is not precluded. This means that infinite and negative values of the safety factor can (according to the mathematics) occur with nonzero probability. As a result, the calculated probability density is not realistic for large deviations from the modal condition. (Even average does not exist).

A Thermodynamic Example (Ascent Heat-Shield Design)

For solid-propellant boosters, an ascent cover may be required to protect the missile during the powered phase of flight. The parameters which affect the design of the cover are the thermal properties and ablation characteristics, the ascent heating, and the characteristics and limitations of the underlying surface. The mechanical properties of the ascent cover material are often (and will, in the following, be assumed to be) such as to impose no restriction on the design. The criterion for success or failure can then be complete ablation of the ascent shield or an allowable temperature (say 900°F) for the undersurface (i.e., the main heat shield of the reentry vehicle, the rocket casing, or the spacer, as the case may be). These two criteria may be identical.

Table 1 is a summary of the thermal properties, ablation characteristics, and the ascent heating for a point on a typical ascent shield, together with the derivatives $\partial t_a/\partial p_i$. Included are nominal values, reference values, and standard deviations. The derivatives were determined by calculating the ablated thickness t_a for the reference conditions, and then, by replacing the reference value of each parameter by its mean value, one at a time, repeating the calculations for t_a, and then, dividing the difference in t_a by the difference $(p_i - p_i^*)$.

Table 2 shows the results of the calculations in terms of the mean and reference ablation, and the standard deviation for four stations on the vehicle based on eqs. (4) and (5). Also listed are reliability estimates for these stations. They are all greater than 0.9999; therefore, if 0.9999 reliability is chosen as the design criterion, the change in thickness required to produce the specified reliability is negative and is also presented in this table. The reference thicknesses for this case were

Table 1. Typical Properties for Ascent Heat Shield

Property	Nominal Values	Reference Values	σ_i	$\left(\frac{\partial t_a}{\partial_{P_i}}\right)$
k	0.110	0.15	0.013	-1.15×10^{-2}
ρ	62.4	60.0	0.8	-5.8×10^{-3}
C_p	0.44	0.42	0.021	-0.226
H_v	600	400	200.0	-3.245×10^{-4}
η_L	0.53	0.51	0.02	-2×10^{-3}
η_T	0.18	0.15	0.03	-.144
ϵ	0.80	0.7	0.10	-4×10^{-4}
T_a	940	840	100.0	-4.07×10^{-4}
Q	4970.0	5670	233.0	$+6.65 \times 10^{-5}$

Table 2. Reliability Estimates for the Ascent Cover

Property	Front of Nose	Rear of Nose	Rear of Mid-section	Front of Flare
t^*	0.60	0.30	0.15	0.30
$(t_a)^*$	0.439	0.223	0.047	0.162
$\sum \frac{\partial t_a}{\partial p_i} (\bar{p}_i - p_i^*)$	-0.169	0.087	0.021	0.063
$\bar{t}_a$	0.27	0.136	0.026	0.099
σ	0.0784	0.039	0.011	0.030
S	3.83	4.205	11.3	6.6
r	0.99993	0.99998	1.0	1.0
$\bar{t} - t^*$	-0.038	-0.019	-0.083	-0.088
$\bar{t}$	0.562	0.281	0.067	0.212

selected on the basis of preliminary work and are seen to be reasonably close to the required thickness, except on the midsection.

The effect of specifying several other probabilities of surviving ascent was investigated and the results are shown in Table 3, which presents the required thickness as a function of reliability for the same stations as Table 2. For this range of reliabilities, approximately 10 per cent increase in weight is required for each additional 9 (decrease by a factor of 10 of the failure probability).

A Thermostructural Example (Re-entry Heat-Shield and Substructure Design)

The design of an ablative-type heat shield for a re-entry vehicle is a typical example of a thermostructural problem in which the proposed method of analysis can be applied. The thermal analysis defines re-entry environment in terms of the ablation (amount ablated) as well as the temperatures of the heat shield and the structure. (See, for instance, Figure 2.) The structural analysis defines the influence of temperature on the structural integrity of the heat shield during re-entry. (See, for instance, Figure 3.) The basis for defining reliability is the critical mode of structural failure. Parametric analyses of a specific vehicle have to be performed to establish the critical mode of failure; for example, a longitudinal crack in the heat shield due to circumferential tension.

Having defined the mode of failure, mathematical models have to be chosen for analyzing failure. It is convenient to set up a single computer program combining thermodynamic equations with structural calculations. The thermal equations consist, essentially, of solving the Fourier heat-conduction equation in one dimension with surface ablation; i.e., with a receding surface. The variables with defined variances involved are re-entry conditions V and γ, conductivity k, density ρ, specific heat C_p, material ablation characteristics designated Q^*, cold-wall heating Q_{cw} and heat-shield thickness t_s. At specified times, the temperature versus thickness output of the thermal calculations can be used in the thermal-stress equations which involve the heat-shield variables; i.e., modulus of elasticity E, coefficient of thermal expansion α, and allowable circumferential tensile stress F.

The thermal-stress equations define the stress in the heat shield using a multi-layered thin-shell analysis of the vehicle. Specifically, the circumferential stress in the i^{th} layer of a multi-layered shell at any time τ is

Table 3. Weight-Reliability Tradeoff

Location	Reliability		
	0.99	0.999	0.9999
Front of Nose	0.453	0.513	0.562
Rear of Nose	0.227	0.257	0.281
Rear of Mid section	0.052	0.066	0.067
Front of Flare	0.170	0.193	0.212
Weight (pounds)	29.0	31.5	33.5

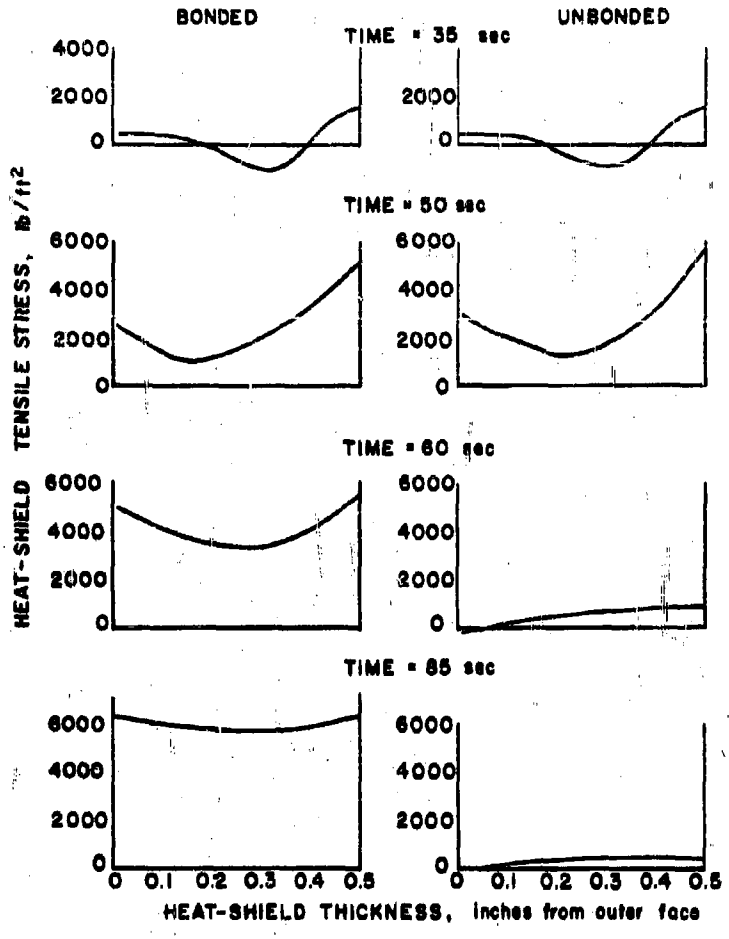


Fig. 3. Heat-Shield Stress Profiles During Re-entry

$$\sigma_i = \left(\frac{E}{1+\nu}\right)_i \left[\frac{\int_0^t \frac{E\alpha T}{1-\nu}\, dx}{\int_0^t \frac{E}{1-\nu}\, dx} - \alpha_i T_i \right] + \frac{(\pi_i - \pi_o)R}{2} \left[\frac{\left(\frac{E}{1-\nu}\right)_i}{\int_0^t \frac{E}{1-\nu}\, dx} \right] \qquad (13)$$

where x is the distance outward from the inside of the shell, and t is the total composite thickness; here T and the pressures are considered to be a function of time, and E and α to be functions of T.

Typical time histories of re-entry conditions are shown in Figure 2, and typical stress profiles at several times during re-entry are shown in Figure 3.

Heat-shield failure is represented by a safety factor of less than 1. Thus, a total of 11 variables are involved in the calculations, where values of each variable have defined variances determined by laboratory test. All variables are considered to be temperature-dependent, but not time-or-load-dependent, and are considered to be statistically independent.

Most properties tend to be in a reasonably Gaussian distribution. If there is a noticeable deviation from normality, it is best not to force a Gaussian distribution with the measured mean and standard deviation through all of the test results, but to fit the results in the area likely to lead to failure; that is, in the region (high or low) of the parameter which aggravates the problem. Such a fit to the allowable tensile strength of an ablation material (in the low-strength region) is shown in Figure 4.

The reference calculation is made using conservative values of all properties, such that the computed safety factor is near unity. The approximate linear tradeoffs (derivatives), mean safety factor, and the standard deviation of the safety factors are computed using Eqs. (4) and (5). Results for a particular station are shown in Table 4. Table 5 shows the reference safety factor, the mean safety factor, and the nominalized safety factor s calculated from Eq. (2); it also shows the standard deviation of the safety factor and, finally, the calculated reliability.

General Remarks on the Small-Perturbations Approach

It has been pointed out that the required partial derivatives have almost always to be obtained by taking the difference of numerically calculated results, varying one of the parameters at a time from a reference condition (and dividing the difference in safety factor by the corresponding difference in the parameter). This raises the question of how much each parameter should be varied. Ideally, a very small difference is the best approximation of what should be a limit as the difference tends to zero. However, very small differences tend to be obscured by machine noise.

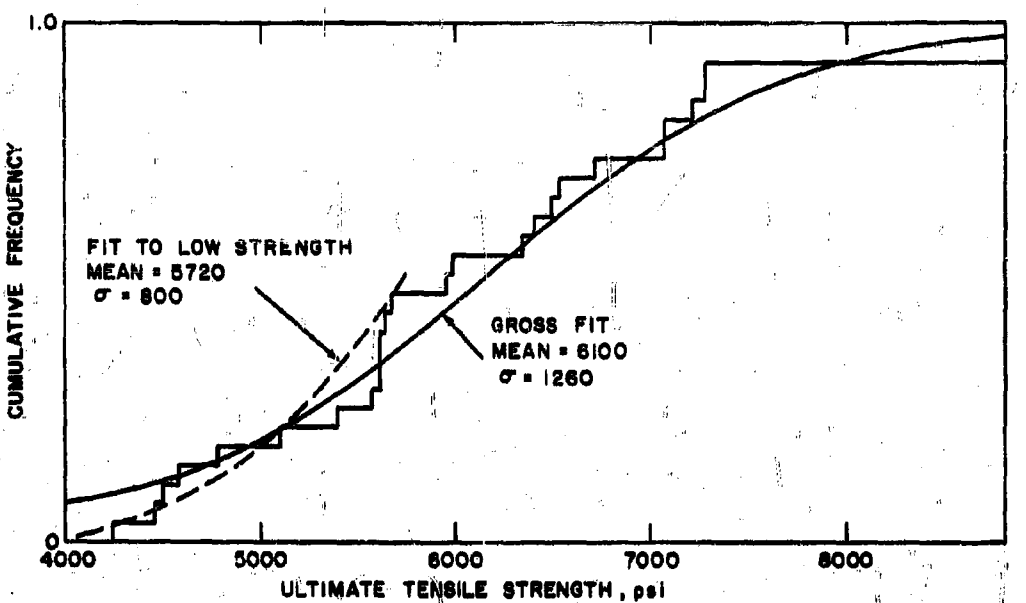


Fig. 4. Allowable Strength of a Typical Ablation Material

Table 4. Typical Heat-Shield Properties

Variable	$\bar{p}$	p^*	σ	$\partial S/\partial p$
E	1.30×10^6	1.53×10^6	0.23×10^6	-0.652×10^{-6}
F	6200	5000	800	0.000224
α	5.07×10^{-6}	4.87×10^{-6}	0.2×10^{-6}	0.11×10^{-6}
V	24,000	24,000	0	0
γ	-19.25	-18.5	0.1667	-0.096
k	0.27	0.305	0.035	-6.1
ρ	95.0	93.7	1.27	0.00935
C_p	0.23	0.21	0.02	5.34
Q^*	4710	4135	1200	0.00014
Q	6660	6660	335	-0.0578
t	0.40	0.40	0.003	6.67

Table 5. Typical Heat-Shield Safety Factor and Reliability

$\sum \frac{\partial S}{\partial p_i} (\bar{p}_i - p_i^*)$	0.9250
S^*	1.120
$\bar{S}$	2.045
σ_S	0.374
s	2.793
r	0.997
$\partial S/\partial W$	0.108
ΔW for $r = 0.999$	+1.04

Therefore, the best guide is that the difference should be the smallest for which the results stand out from the machine noise, and generally, not larger than a 1σ variation in the given parameter.

The possibility usually exists of programming a linearized small-perturbation version of the given mathematical model, and then, solving for the derivatives directly. The procedure is to write the model first for an arbitrary set of parameters, then for a set of parameters all disturbed by a small amount from the first set, expand all functions in the second case involving the perturbed parameters in Taylor series, retaining only the linear terms, to subtract from this set of equations the set of equations for the unperturbed case. The result is a set of linear (generally ordinary or partial differential) equations for the desired derivatives; these equations involve the solution of the given mathematical model for the unperturbed case.

This approach can be illustrated very sketchily by means of the simple mathematical model of Eq. (8). (This illustration misses, unfortunately, many of the subtle problems involved in this approach when used in conjunction with complicated sets of nonlinear differential equations.) Let the perturbed variables be $S + \delta S$, $E + \delta E$, etc., so that Eq. (8) can be written as

$$S + \delta S = \frac{K(E + \delta E)}{1 - \nu^2} \frac{(t + \delta t)^3}{b(P + \delta P)}$$

and expanding in Taylor series,

$$S + \delta S = \frac{KE}{(1 - \nu^2)} \frac{t^3}{bP} + \frac{K\delta E}{(1 - \nu^2)} \frac{t^3}{bP} + \frac{KE\,3t^2}{(1 - \nu^2)bP} \delta t - \frac{KE\,t\,\delta P}{(1 - \nu^2)bP^2}$$

so that, upon subtracting Eq. (8),

$$\delta S = S \left[\frac{\delta E}{E} + 3 \frac{\delta t}{t} - \frac{\delta P}{P} \right]$$

which then gives, directly, the derivatives given in Eqs. (10), (11), and (12).

The formal small-perturbation mathematical-model approach leads to an increased accuracy in the computation, and possibly, a saving in computing time. However, it requires a very sophisticated applied-mathematical analysis and high-level programming, and it is not clear that the benefits warrant such an effort.

One form of nonlinearity in materials properties that occurs

occasionally is a very abrupt change, almost a discontinuity, with one of the parameters of concern, say, temperature (see, for instance, Figure 5.) One way to treat this problem by means of a linearized approach is to make two analyses using reference conditions which differ only in the two different values of the parameter (one on each side of the discontinuity) causing the nonlinearity; by determining the conditional probability that the given parameter falls above or below the critical value, and then, putting the two results for the statistics of the safety factor together using this conditional probability, one can get the composite statistics for the safety factor.

It may be observed here that if the phenomenon is nonlinear, the quantities $\bar{S}$ and σ_S calculated from a reference condition have no direct significance; they are not the true mean and standard deviation of S, but merely quantities which, when substituted in the relation for a Gaussian probability distribution, give a close approximation to the true probability of S in the region corresponding to failure (much as the fit used in Figure 4 to approximate the critical tail end of the distribution leads to a fictitious mean value and standard deviation). This phenomenon is illustrated (in Figure 6) for a typical variation of S with a single parameter. These remarks are made here because, occasionally, the true S can be estimated independently, and the S calculated by the linearized approach is, then, found to disagree with it; from what has been said here, it is obvious that close agreement should not always be expected.

For design purposes, it is useful to know the change of thickness required or the weight penalty involved in attaining a desired reliability. Again using the assumption of linearity, the safety factor for any heat-shield weight W can be expressed in terms of a weight change; namely,

$$S = S^* + \sum_{i=1}^{n} \left(\frac{\partial S}{\partial p_i}\right)^* (p_i - p_i^*) + \left(\frac{\partial S}{\partial W}\right)^* (W - W^*) \qquad (14)$$

Consequently, if a given reference design (with reference weight W^*) has a value of S, implying an unsatisfactory reliability, one can obtain a new value of S from the desired reliability, and a new weight from the preceding equation.

That is, with the mean value of S

$$W = W^* + (S_{new} - S) \frac{\sigma_S}{\left(\frac{\partial S}{\partial W}\right)^*} \qquad (15)$$

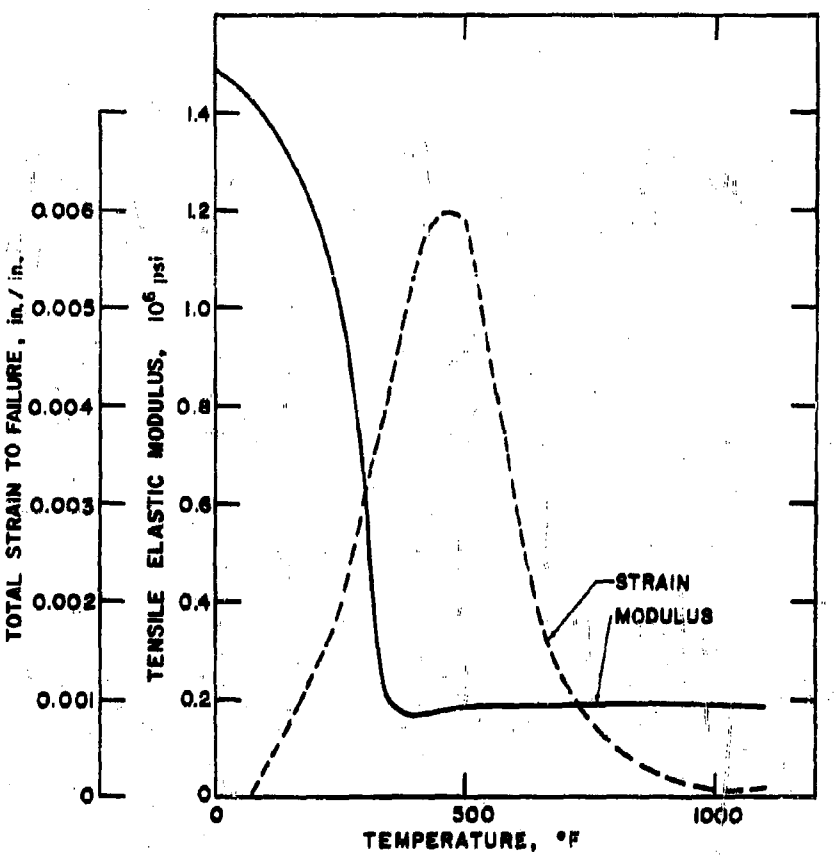


Fig. 5. Properties of a Typical Ablation Material

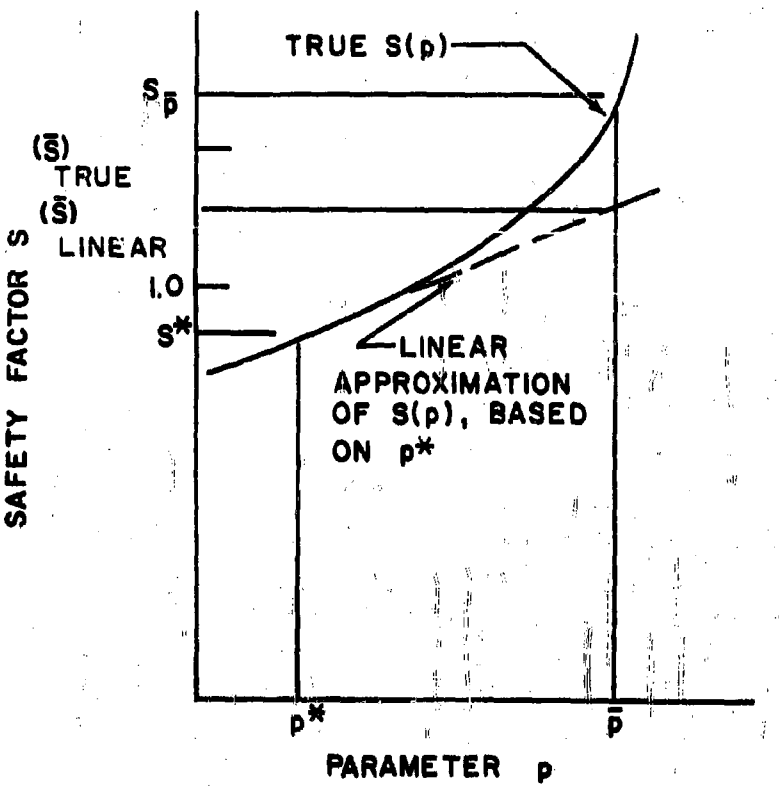


Fig. 6. Nonlinear Variation of the Safety Factor S with a Single Parameter

THE MONTE-CARLO APPROACH

When the effect of nonlinearities in the dependence of the safety factor on the mechanical and thermal properties, or the applied heating and loading are too large to be treated even with a suitably chosen reference condition, the most practical alternative is to resort to a Monte-Carlo procedure.

To do so requires that a mathematical model for the safety factor in terms of the several input variables exists, and that the statistics of the variables be known; they need not be Gaussian or independent. The technique consists in selecting sets of input variables at random, but in such a way that the statistics of the selections represent the actual statistics.

If the probability densities of the (say, n) variables are independent and Gaussian, it is necessary only to select sets of n numbers from a table of Gaussian random numbers (1), multiply each of the n numbers by the standard deviation of one of the n variables, and add to this the mean value of the corresponding variable. The resulting n values for the n variables, then, represent the input for a given calculation of the safety factor.

If the variables are Gaussian but are correlated, one can use a transformation technique. Let, for instance, the j^{th} and k^{th} variables be correlated, with the correlation coefficient ψ_{jk}. For this pair of variables, two numbers n_j and n_k are chosen as before for each calculation out of a table of Gaussian random numbers, but instead of multiplying these by σ_j and σ_k, respectively, they are subjected to the following transformation:

$$\begin{bmatrix} P_j \\ P_k \end{bmatrix} = \begin{bmatrix} \bar{P}_j \\ \bar{P}_k \end{bmatrix} + \frac{\sqrt{2}}{2} \begin{bmatrix} \sqrt{1+\psi_{jk}}\,\sigma_{P_j} & -\sqrt{1-\psi_{jk}}\,\sigma_{P_j} \\ \sqrt{1-\psi_{jk}}\,\sigma_{P_k} & \sqrt{1-\psi_{jk}}\,\sigma_{P_k} \end{bmatrix} \begin{bmatrix} n_j \\ n_k \end{bmatrix} \tag{16}$$

If the variables are known to be non-Gaussian, two techniques can be used. The best consists in using the actual test data which were used to define the statistics. Suppose a given property was defined by a series of N measurements, which when analyzed indicate that the probability density is non-Gaussian. For the purpose of the safety-factor calculations, one can, in each calculation, choose one of the N measurements at random, making sure that each measurement has an equal chance of being selected.

table of the area under the normal curve.

In some cases, a nonlinear problem can be resolved into two phases; i.e., linear and nonlinear; and sometimes the nonlinear phase can be handled by itself by means of a very simple mathematical model, most of the complication residing in the linear phase. One such example that has been encountered is the following.

The critical mode of failure of a certain heat-shield/structure composite was determined to be a strain failure occurring at a time near impact, when the heat shield and structure were very nearly at a uniform temperature. Since temperature at impact varied linearly with thermal properties, it was possible to treat the heat-shield temperature as a linear function of the thermal parameters and, therefore, to establish a mean and standard deviation of the temperature distribution. Thus, the linearized-perturbation approach could be used to establish the statistics of the temperature, an intermediate result, and a Monte-Carlo approach, then is used to obtain the statistics of the safety factor in terms of temperature and the mechanical properties. In this manner, the very sophisticated and time-consuming ablation calculations had to be performed only often enough to establish the required derivatives, whereas the very simple thermal-stress calculations could readily be performed very many times in a short time (about 1000 calculations per minute on a Philco S2000 computer).

The results of this particular analysis showed that the safety factor was indeed not normally distributed, and that a large percentage of the cases produced very large safety factors. The cumulative distribution is shown (in Figure 7) for the first 28 cases out of 1000, while the frequency distribution for all the cases is shown in Figure 8. Estimations of the heat-shield reliability were made by curve fitting Figure 7 in the region of interest; namely, near $S = 1$. Upper and lower estimates of the "best" fit showed the calculated reliability to be between 0.9985 and 0.9997.

CONFIDENCE LIMITS

As mentioned in the introduction, one purpose of a safety factor is to cover uncertainties. The approach outlined in this paper is based on a probabilistic approach to the establishment of safety factors for designing hardware; consequently, it addresses itself primarily to the random factors. Nonetheless, recognition must be given to the as-yet-unresolved very basic problem of the relation between uncertainty and randomness. It is not the purpose herein to contribute any new insights into this problem, but to effect the required synthesis of uncertainty and randomness, uncertainties will be related to a known concept; namely, confidence limits,

One way of doing this is to list the N measurements in a table opposite an index a, with $a = 1, 2, \ldots, N$, then in each calculation, choose an a from a table of (uniform) random numbers, and next, looking up the corresponding measurement from the table. An alternative consists in finding some analytical function $p' = f(p)$, of the given property, such that p' is normally distributed when p is distributed in the given manner. Then, choosing Gaussian random values for p' and applying the inverse function, yields properly distributed values of p. This is, in general, the less practical approach. For non-Gaussian correlated variables, it is best to use the actual test data.

The problem of choosing input values for the several variables is, therefore, a fairly trivial one. The practical difficulty is in making large numbers of calculations by means of a sophisticated mathematical model, which requires large amounts of computing time for each calculation. The only solution is to use as simple a model and as few calculations as possible, while still retaining the desired accuracy and confidence limits.

The required number of calculations is related to the confidence limits of the results, as determined by the sample sizes of the input parameters. The number of the calculations should be such as not to degrade significantly the quality of information contained in the inputs. A conservative criterion is three times the largest of the input sample sizes. Allowing for the relative influence of the several parameters on the safety factor, the criterion can be modified to three times the sample size for which

$$N_i \Big/ \left(\frac{\partial S}{\partial p_i}\right)^2 \sigma_i^2$$

is largest. This derivative is not normally obtained in a Monte-Carlo calculation, but for the purposes of this criterion only a qualitative measure is required, and this can often be deduced from the calculations.

The result of these calculations is, then, a probability density of the safety factor, for which the fraction of cases falling above $S = 1$ must be deduced. The obvious approach is one of enumerating this fraction directly by dividing the number of cases for which $S > 1$ by the total number. Unless the total number is very large, this directly calculated number is not likely to furnish reliability estimates with high confidence, particularly if a very high reliability is the design criterion. More sophisticated methods of extreme-value statistics must then be used (2). If the distribution is nearly Gaussian, this can be done readily by means of the calculated average and standard deviation (defined over the given sample size), using a

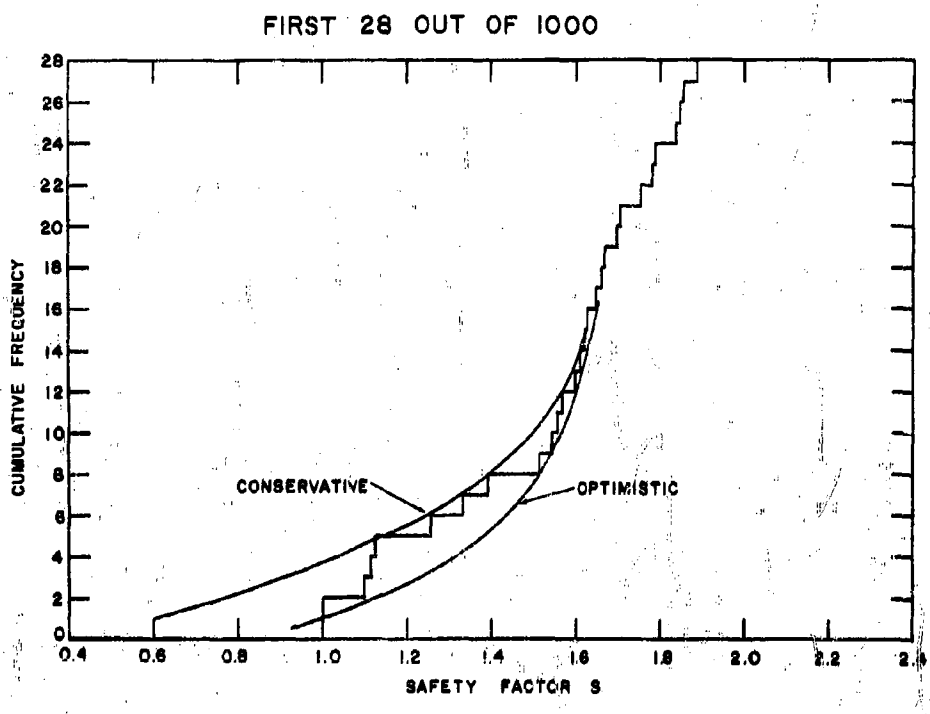


Fig. 7. Cumulative Safety Factors

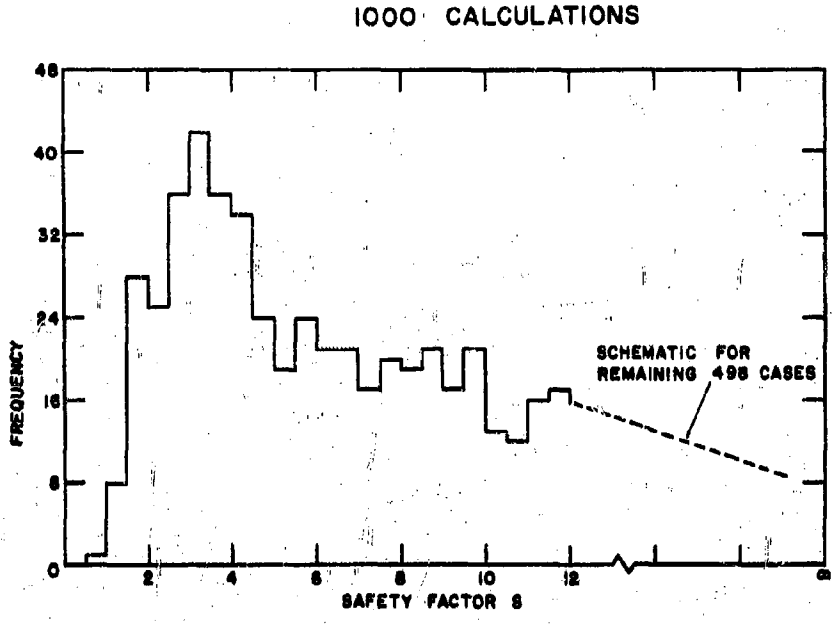


Fig. 8. Frequency Distribution of Safety Factor

use of which in conjunction with statistical estimates is well understood.

The first problem to be considered is a basic uncertainty in the measurement of a quantity (say, due to limitations in the apparatus). It is shown in Appendix B that even if very many ($N \to \infty$) determinations are made, the apparatus uncertainty may be interpreted as resulting in an implied sample size N; (the larger the uncertainty, the smaller the implied sample size). Suppose the uncertainty is a given fraction of the mean, so that all measurements of the quantity p_i are really $p_i \pm \epsilon \bar{p}_i$, then, $\hat{N}$ can be calculated as a function of the parameter $(\sigma_{p_i}/\bar{p}_i)$. A short table (for $\sigma_{p_i}/\bar{p}_i = 0.05$ and for two confidence level) is given in Table 6.

Next, the combination of the implied with the actual finite sample size has to be considered; suppose a property is measured N times with an apparatus of a given uncertainty ϵ. The implied sample size must now certainly be less than N or $\hat{N}$, whichever is smaller. A rigorous analysis cannot be performed because a random and a nonrandom, but partially unknown, process are involved. In analogy with the relation below for the effective sample size of S in terms of the sample sizes of the parameters, the following relation suggests itself:

$$N_{\text{combined}} = \frac{1}{\frac{1}{\hat{N}} + \frac{1}{N}} \tag{17}$$

The third problem is to estimate the uncertainty of the output S in terms of the uncertainties in the inputs (the various properties and environmental parameters). For the uncertainty in any one parameter, an uncertainty in S can be calculated directly from the actual or linearized mathematical model. The question is how to combine the contributions from the several parameters; clearly; they cannot be root-mean-squared since they do not represent random factors; nor does it appear to be reasonable to add them arithmetically.

The sample-size concept furnishes a means of effecting this combination. In the case of the linearized approach, an equivalent sample size for S can be obtained readily in terms of the sample sizes for the parameters.

Table 6. Implied Sample Size

(For Infinite Actual Sample Size; $\sigma_i/\bar{p}_i = 0.05$)

Uncertainty ϵ (As a Fraction of Mean Value)	$\hat{N}$ for 90 percent Confidence	$\hat{N}$ for 75 percent Confidence
0.01	70	43
0.02	18	12
0.03	9	6
0.04	6	4
0.05	5	3

$$N_s = \frac{1}{\sum_{i=1}^{n} \left(\frac{\partial S}{\partial_{P_i}}\right)^2 \frac{\sigma_{P_i}^2}{\sigma_s^2} \frac{1}{N_1}} \tag{18}$$

No such simple analysis can be used in conjunction with the Monte-Carlo approach. (Certainly, it would be a fallacy to use the number of runs as an indication of confidence because that number may be very large, whereas some of the input data may have been obtained with a small actual or implied sample size; the use of these data again and again in the process of making the random computer runs does not, of course, improve the inherent accuracy of the data.) In the absence of a rigorous criterion, Eq. (18) may be used as a rough guide to the confidence that can be attached to the results of the Monte-Carlo analysis; the derivatives can be obtained by numerical differencing, as indicated previously.

Once the implied sample size of S has been established, it can be used to assign confidence limits to the reliability calculations in the conventional manner, using, for instance, the non-central t distribution (3).

EXTENSION TO MULTI-ELEMENT SYSTEMS

The preceding discussion pertains, essentially, to a single element and its probability of survival. The problem remains of combining these analyses for a number of elements into a composite thermostructural system reliability by means of a system failure-mode analysis. Clearly, it is incorrect to treat all elements as being independent because both the environments and the properties of various elements bear a certain relation.

In the case of the heat shield covering an entire re-entry vehicle; for instance, the aerodynamic heating will differ at different points on the vehicle, but the statistical variation of the heating, which is trajectory-dependent, is the same for the entire heat shield. The materials properties will tend to vary between different segments and, possible if the segments are large and widely separated, points are considered within the segments. (At closely spaced points within segments, the properties will, of course, tend to be the same.) Therefore, the properties can be considered independent if relatively few points are considered. A composite failure-mode analysis can then be made on the basis of this hypothesis.

As more and more complex systems are analyzed, and more and more points on the vehicle, as well as more and more environments have to be considered, the greater sophistication of this approach over the more conventional one may become economically untenable. In this event, a different philosophy can be used which has been used for aircraft for several decades (4). This consists in making the more sophisticated analysis at a few points and conditions, expressing the results in an effective safety factor based on a relevant, and simple analyzable, failure mode and criterion, and use the latter for the entire vehicle. Thus, essentially, reliability considerations are used to define a safety factor, and the vehicle is designed for that safety factor.

Since the present approach also affords a means of calculating thermostructural reliabilities, apart from any considerations of design criteria, it may be pertinent to remark that it has applications in the field of reliability apportionment. At present, the reliability is often apportioned to the subsystems in a somewhat haphazard way. The desirable approach is, of course, to allocate it on a systems analytical basis, in such a way that the system performance (however defined, including cost and schedule consideration if one wishes) is maximized.

A great deal of work has been done in this field of structural optimization for maximum performance at a specified reliability (5) and (6). The approach outlined herein forms a useful and almost essential ingredient of such analyses. In particular, the partial derivatives used in the small-perturbation approach offer a very convenient tool to use in conjunction with dynamic-and-linear-programming techniques, as well as a means of assessing intuitively where the greatest opportunities for improving performance lie. It is expected, therefore, that the approach of the present paper will find many applications along these lines.

CONCLUDING REMARKS

A method has been outlined for using a specified level of reliability as a design criterion for thermostructural and similar complex systems subject to a combination of environments as an alternative to the conventional approach of using combinations of safety factors for several environments. This approach requires the efficient calculation of the reliability of such systems, and two methods for doing so have been discussed at some length; i.e., a linearized-perturbation method (with special techniques to adapt it to nonlinear problems) and a Monte-Carlo method.

The acquisition of the required input statistics and the use of the methods in problems of practical interest have been discussed in some detail, and several examples have been cited. The role of uncertainties (as opposed to radom variations) has been discussed, and an implied-sample-size concept has been introduced, whereby these uncertainties can be translated into confidence estimates for the reliabilities caculated by this approach. The use of this approach in conjunction with failure-mode analyses of an optimal reliability apportionment to thermostructural systems has been indicative.

It is felt that the approach offered herein has borne fruit already in the efficient design of several heat-shield/substructure composites of re-entry vehicles, and that it represents a powerful tool for the solution a wide range of design and systems analytical problems of sophisticated systems subjected to complex combinations of environments.

APPENDIX A. ADDITIONAL STATISTICAL CONSIDERATIONS

The normal (Gaussian) probability-density function,

$$G(p) = \frac{1}{\sqrt{2\pi}\sigma_p} e^{-\frac{1}{2}\left(\frac{p-\bar{p}}{\sigma_p}\right)^2} = \frac{1}{\sigma_p} \phi\left(\frac{p-\bar{p}}{\sigma_p}\right)$$

describes the statistics of many physical processes, largely because, if a physical random process is itself the result of many independent random processes, its probability density will be Gaussian. Thus, if a material property is the result of the interaction of many minute elements in the material, it will tend to be Gaussian, within certain limits. However, if beyond a certain extent the dependence of the property on the number of "flaws" is nonlinear then the property will not be Gaussian for large deviations of the property from its mean.

Consequently, although the Gaussian distribution will often fit data very well in the region in the vicinity of the mean (2 or 3 σ on either side), it will not generally fit the data in the region of large differences from the mean. (The probability of getting negative mechanical and thermal properties for the usual materials is zero for instance, although the Gaussian distribution predicts a small probability of incurring them.)

The intergrated probability density; that is, the distribution function for the Gaussian case is given by

$$G(p) = \int_{-\infty}^{p} g(p')\,dp' = \Phi\left(\frac{p-\bar{p}}{\sigma_p}\right)$$

Here, ϕ and Φ are the normalized Gaussian probability density and distribution function ("area under the normal curve"), respectively, so that

$$\Phi(x) = \int_{-\infty}^{x} \phi(x')\,dx'$$

This function is tabulated in (2) and elsewhere although most tabulations lack the required number of significant figures. (The value 0.500000 must be added to the numbers listed in Table II (2) to obtain Φ as defined herein.) An asymptotic expression for Φ for large values of its argument is

$$\Phi(x) \approx 1 - 0.3989\,\frac{e^{-\frac{x^2}{2}}}{x}\left[1 - \frac{1}{x^2} + \frac{1.3}{x^4} - \ldots -\right]$$

If the specific variable of interest is the safety factor, and it is so defined that failure will occur for $S < 1$, then, the reliability is

$$r = \text{Prob}\{S > 1\}$$

$$= \int_1^{\infty} [\text{probability density of } S]\, dS$$

$$= 1 - \int_0^1 [\text{probability density of } S]\, dS$$

and if S is Gaussian,

$$r = 1 - \Phi(-s)$$

and

$$= \Phi(s)$$

where s is defined in Eq. (2).

The basic statistical problem herein is the determination of the statistical characteristics of S (whether they turn out to be Gaussian or not) from those of various input parameters (properties and environmental characteristics). The direct statistical approach is the following.

Given a functional relationship which relates S to the various parameters,

$$S = \tilde{S}(p_i), \qquad i = 1, 2, \ldots, n$$

and the joint distribution function of the n variables

$$F(X_i) = \text{Prob}\left\{p_i \leq X_i, \qquad i = 1, 2, \ldots, n\right\}$$

$$f(X_i) = \frac{\partial^n F(X_i)}{\partial X_1 \partial X_2, \ldots, \partial X_n}$$

then, the probability density of S is

$$g(S) = \underbrace{\int\int\int \cdots \int}_{\tilde{S}(p_i) = S} \frac{f(X_i)}{|\text{grad } \tilde{S}|} \, dX_1 \, dX_2, \ldots, dX_n$$

that is,

if the n variables are all independent., f is a product distribution.

$$f(X_i) = f_1(X_1) f_2(X_2) \ldots f_n(X_n)$$

This direct approach may be referred to as the convolution method, and generally is not very practical in the problems of concern here. A very simple example is the case where S depends on only two parameters f and F ; namely,

$$S = \frac{F}{f}$$

and both f and F are Gaussian. The distribution of S can readily be calculated and is

$$g(S) = \frac{1}{\sqrt{2\pi}} \frac{\bar{f} \, \sigma_F^2 + S \bar{F} \, \sigma_f^2}{\left\{\sigma_F^2 + S^2 \, \sigma_f^2\right\}^{3/2}} e^{-1/2 \frac{(S\bar{f} - \bar{F})^2}{\sigma_F^2 + S^2 \sigma_f^2}}$$

However, this distribution has no (finite) mean or standard deviation as a result of the fact that for a Gaussian distribution, f has a nonzero probability of being zero, so that S may be infinite (or negative) with significant probability.

If it is now stipulated that the standard deviations of f and F are small compared to the mean values, and if the distribution of f and F are assumed to be trunacted at $n\sigma$ on either side of the mean then, it will be found that

$$\bar{S} = \frac{\bar{F}}{\bar{f}} + \epsilon_{\bar{S}}$$

and $$\sigma_S^2 = \frac{\sigma_F^2 \bar{f}^2 + \bar{F}^2 \sigma_f^2}{\bar{f}^4} + \epsilon_\sigma$$

where the truncation errors can be specified by absolute upper bounds; namely,

$$\left| \epsilon_{\bar{S}} \right| \leq \frac{\bar{F}}{\bar{f}} \left(\frac{n \sigma_f}{\bar{f}} \right)^2$$

and $$\left| \epsilon_\sigma \right| \leq \frac{3\sigma_F^2 + \bar{F}^2}{\bar{f}^2} \left(\frac{n \sigma_f}{\bar{f}} \right)^2$$

For most cases, this direct convolution approach is not practical. The linearized-perturbation approach is capable of handling many cases of interest, particularly, if nonlinearities are taken into account approximately by the choice of a proper reference condition. The derivation of the required equations is quite straightforward. A mathematical model can be constructed in linearized form to include S and the parameters of interest. This contains

$$S = S^* + \sum_{i=1}^{n} \left(\frac{\partial S}{\partial p_i} \right)^* (p_i - p_i^*)$$

and the definitions of the statistical characteristics

$$\bar{S} = \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} S_j$$

and

$$\sigma_S^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} (S_j - \bar{\bar{S}})^2$$

where the sums are now over a sample of size N (as $N\to\infty$), and j designates the j^{th} element in the sample. Substituting the linearized equation for S into the definition, inverting the order of summation, and making use of the definitions of the mean values and standard deviations of the parameters p_i which are similar in form to that for S , one obtains Eqs. (4) and (5), provided that p_i is uncorrelated.

Last, it may be pertinent to comment on the relation between normality and linearity. A variable will be Gaussian if it is a linear function of a number of parameters which are Gaussian (such as S in the linearized model). If the variable is a linear function of very many non-Gaussian parameters which are independent and of the same general magnitude; that is, if the terms $(\partial S / \partial_{p_i})\,\sigma_{p_i}$ are of the same order of magnitude, the variable should have an approximately Gaussian distribution. (However, if a variable is a heavily nonlinear function of one or a few parameters (as S usually is), even if the parameters are Gaussian and independent, the variable itself will be non-Gaussian. Thus, in general, S will not be Gaussian although in many cases it may not deviate greatly from Gaussian for several standard deviations. Thus the Gaussian expressions can be used for reliability estimates.

APPENDIX B. IMPLIED SAMPLE SIZE

Given an estimate of the mean $\bar{x}$ of a Gaussian random variable with standard deviation $\hat{\sigma}$, it is assumed that $\bar{x}$ is known with an accuracy of ϵ ; that is, $\pm \epsilon\bar{x}$ represents the limits of an unknown bias, so that $x \pm \epsilon\bar{x}$ represents an interval within which the true mean is expected to lie with some confidence. This confidence is largely psychological; i.e., it depends on experience and intuition rather than on quantitative analysis, and certainly is not based on any probabilistic considerations.

The statistical counterpart of this engineering concept is the confidence interval

$$\bar{x} \pm \frac{t_{\gamma, N-1}}{\sqrt{N}} \hat{\sigma} \tag{B-1}$$

where $t_{\gamma, N-1}$ is the 100 γ percentage point of the Student t distribution with N-1 degrees of freedom. (See, for instance (7), wherein $t_{\gamma, N-1}$ is tabulated as a function of $0.5 - \gamma$.) This confidence is based on well defined probabilistic considerations.

It is desired to estimate the probability that the random variable x will exceed a lower specification limit X; that is, to estimate the probability

$$r = P\{x > X\} \tag{B-2}$$

with confidence level 2γ. To make this estimate, it is necessary to have the sample size N associated with $\bar{x}$ and $\hat{\sigma}$. For this purpose, equate the two bracket widths so that

$$\epsilon_{\bar{x}} = \frac{t_{\gamma, N-1}}{\sqrt{N}} \hat{\sigma} \tag{B-3}$$

and solving for N gives

$$N = K t^2_{\gamma, N-1} \tag{B-4}$$

where the constant K is defined by

$$K \equiv \left(\frac{\hat{\sigma}}{\epsilon_{\bar{x}}} \right)^2$$

Equation (B-4) can be solved by an iterative procedure for a given value of K and a given confidence level 2γ. The value of N obtained in this manner is called the implied sample size $\hat{N}$. It thus represents the amount of information or experience implied by the estimate of the uncertainty ϵ.

REFERENCES

1. Rand Corporation, A Million Random Digits with 100,000 Normal Deviates, The Free Press Publishers, Glencoe, Illinois, 1955.

2. Hoel, P. G., Introduction to Mathematical Statistics, John Wiley and Sons, New York, 1947.

3. Johnson, N. L. and B. L. Welch, "Applications of the Non-Central t Distribution," Biometrika, Vol. 31, 1939, pp. 362-389.

4. Bouton, I., "The Fundamental Aspects of Structural Reliability," IAS Paper No. 62-32, January 22-24, 1962.

5. Hilton, H. and M. Feigen, "Minimum Weight Analysis Based on Structural Reliability," Journal of the Aerospace Sciences, vol. 27, no. 9, September 1960, pp. 641-652.

6. English, J. M., "The Effects of Tolerances upon Economy and Safety of Structures, "Boeing Airplane Company, Document D-14589, September 1954.

7. Bowker, A. H. and G. J. Lieberman, Engineering Statistics, Prentice-Hall, Englewood Cliffs, N.J., 1959.